AD-A119 240   BOSTON UNIV MA CENTER FOR ADAPTIVE SYSTEMS   F/6 5/10
                THE QUANTIZED GEOMETRY OF VISUAL SPACE: THE COHERENT COMPUTATIO--ETC(U)
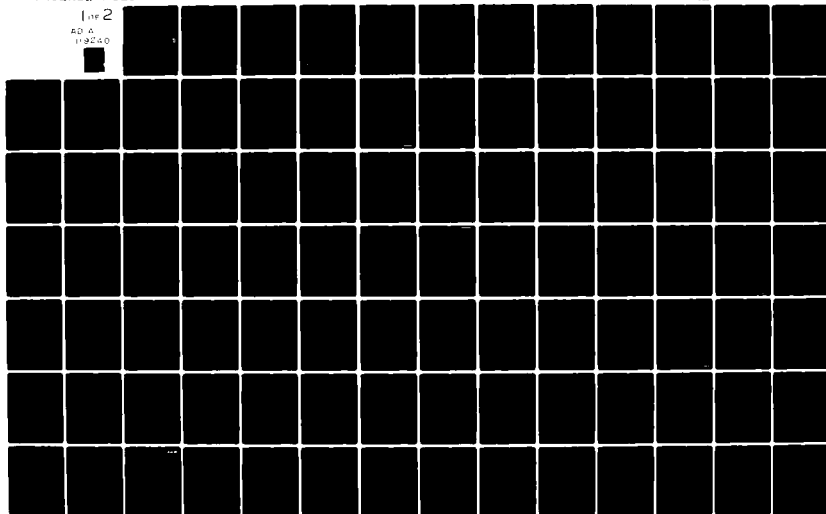   AUG 82   S GROSSBERG   AFOSR-82-0148
UNCLASSIFIED   AFOSR-TR-82-0713   NL

1 OF 2

AD A
119240

AFOSR-TR. 82-0713 ④

THE QUANTIZED GEOMETRY OF VISUAL SPACE:

THE COHERENT COMPUTATION OF

DEPTH, FORM, AND LIGHTNESS

Revised Version

by

Stephen Grossberg*

Center for Adaptive Systems

Department of Mathematics

Boston University

Boston, Mass.   02215

AD A119240

DTIC FILE COPY

TABLE OF CONTENTS

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 82-0713	**2. GOVT ACCESSION NO.** AD-A119240	**3. RECIPIENT'S CATALOG NUMBER**
4. TITLE (and Subtitle) THE QUANTIZED GEOMETRY OF VISUAL SPACE: THE COHERENT COMPUTATION OF DEPTH, FORM, AND LIGHTNESS		**5. TYPE OF REPORT & PERIOD COVERED** Interim Report 01/15/82 - 01/14/83
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Stephen Grossberg		**8. CONTRACT OR GRANT NUMBER(s)** AFOSR 82-0148
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Adaptive Systems Mathematics Department, Boston University Boston, MA 02215		**10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS** 61102F 2313/A5
11. CONTROLLING OFFICE NAME AND ADDRESS AFOSR - Life Sciences Directorate Bolling Air Force Base Washington, DC 20332		**12. REPORT DATE** August, 1982
		13. NUMBER OF PAGES 115 pages
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		**15. SECURITY CLASS. (of this report)** unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release;
distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

binocular vision, neural networks, lightness perception, form perception,
contrast enhancement, filling-in, receptive fields, visual scaling,
resonant neural activity

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A theory is presented of how global visual interactions between depth, length,
lightness, and form percepts can occur. The theory suggests how quantized activity
patterns which reflect these visual properties can coherently fill-in, or complete,
visually ambiguous regions starting with visually informative data features. Pheno-
mena such as the Cornsweet and Craik-O'Brien effects, phantoms and subjective con-

20. ABSTRACT

tours, binocular brightness summation, the equidistance tendency, Emmert's law, allelotropia, multiple spatial frequency scaling and edge detection, figure-ground completion, co-existence of depth and binocular rivalry, reflectance rivalry, Fechner's paradox, decrease of threshold contrast with increased number of cycles in a grating pattern, hysteresis, adaptation level tuning, Weber law modulation, shift of sensitivity with background luminance, and the finite capacity of visual short term memory are discussed in terms of a small set of concepts and mechanisms. Limitations of alternative visual theories which depend on Fourier analysis, Laplacians, zero-crossings, and cooperative depth planes are described. Relationships between monocular and binocular processing of the same visual patterns are noted, and a shift in emphasis from edge and disparity computations towards the characterization of resonant activity-scaling correlations across multiple spatial scales is recommended. This recommendation follows from the theory's distinction between the concept of a structural spatial scale, which is determined by local receptive field properties, and a functional spatial scale, which is defined by the interaction between global properties of a visual scene and the network as a whole. Functional spatial scales, but not structural spatial scales, embody the quantization of network activity that reflects a scene's global visual representation. A functional scale is generated by a filling-in resonant exchange, or FIRE, which can be ignited by an exchange of feedback signals among the binocular cells where monocular patterns are binocularly matched.

Accession For

NTIS GRA&I

DTIC TAB

Unannounced

Justification

By

Distribution/

Availability Codes

Dist | Avail and/or

Special

A

TABLE OF CONTENTS

-Cont'd-

Page

PART II

TABLE OF CONTENTS

-Cont'd-

ABSTRACT

A theory is presented of how global visual interactions between depth, length, lightness, and form percepts can occur. The theory suggests how quantized activity patterns which reflect these visual properties can coherently fill-in, or complete, visually ambiguous regions starting with visually informative data features. Phenomena such as the Cornsweet and Craik-O'Brien effects, phantoms and subjective contours, binocular brightness summation, the equidistance tendency, Emmert's law, allelotropia, multiple spatial frequency scaling and edge detection, figure-ground completion, coexistence of depth and binocular rivalry, reflectance rivalry, Fechner's paradox, decrease of threshold contrast with increased number of cycles in a grating pattern, hysteresis, adaptation level tuning, Weber law modulation, shift of sensitivity with background luminance, and the finite capacity of visual short term memory are discussed in terms of a small set of concepts and mechanisms. Limitations of alternative visual theories which depend upon Fourier analysis, Laplacians, zero-crossings, and cooperative depth planes are described. Relationships between monocular and binocular processing of the same visual patterns are noted, and a shift in emphasis from edge and disparity computations towards the characterization of resonant activity-scaling correlations across multiple spatial scales is recommended. This recommendation follows from the theory's distinction between the concept of a structural spatial scale, which is determined by local receptive field properties, and a functional spatial scale, which is defined by the interaction between global properties of a visual scene and the network as a whole. Functional spatial scales, but not structural spatial scales, embody the quantization of network activity that reflects a scene's global visual representation. A functional scale is generated by a filling-in resonant exchange, or FIRE, which can be ignited by an exchange of feedback signals among the binocular cells where monocular patterns are binocularly matched.

The objects of perception and the space in which they seem to lie are not abstracted by a rigid metric but a far looser one than any philosopher ever proposed or any psychologist dreamed.

Jerome Lettvin (1981)

1. Introduction: The Abundance of Visual Models

Few areas of science can boast such a wealth of interesting and paradoxical phenomena that are readily accessible to introspection as visual perception. The sheer variety of effects helps to explain why so many different types of theories have arisen to carve up this data landscape. Fourier analysis (Cornsweet, 1970; Graham, 1981; Robson, 1975), projective geometry (Beck, 1972; Johannson, 1978; Kaufman, 1974), Riemannian geometry (Blank, 1978; Luneberg, 1947; Watson, 1978), special relativity (Caelli, Hoffman, and Lindman, 1978), vector analysis (Johannson, 1978), analytic function theory (Schwartz, 1980), potential theory (Sperling, 1970), cooperative and competitive networks (Amari and Arbib, 1977; Dev, 1975; Ellias and Grossberg, 1975; Grossberg, 1970, 1973, 1978a, 1981; Sperling, 1970; Sperling and Sondhi, 1968) are just some of the formalisms which have been used to interpret and explain particular visual effects. Some of the most distinguished visual researchers believe that this diversity of formalisms is inherent in the nature of psychological phenomena. Sperling (1981) has, for example, recently written:

"In fact, as many kinds of mathematics seem to be applied to perception as there are problems in perception. I believe this mul-

tiplicity of theories without a reduction to a common core is in-

herent in the nature of psychology..., and we should not expect

the situation to change. The moral, alas, is that we need many

different models to deal with the many different aspects of per-

ception" (p. 282).

The opinion which Sperling offers is worthy of the most serious deliberation, since

it predicts the type of mature science which psychology can hope to become, and there-

by constrains the type of theorizing which psychologists will try to do. Is Sperling

correct, or do there exist concepts and properties, heretofore not explicitly incor-

porated into the mainstream visual theories, which can better unify the many visual

models into an integrated visual theory?

Part I of this article reviews various visual data as well as internal paradoxes and inherent limitations of some recent theories that have attempted to explain these data. Part II presents a possible approach to overcoming these paradoxes and limitations and to explaining the data in a unified fashion. The two parts of the paper are self-contained and can be read in either order.

<div align="center">PART I</div>

2. The Quantized Geometry of Visual Space

There is an important sense in which Sperling's assertion is surely true, but this sense is shared with other sciences such as physics. Different formalisms can probe different levels of the same underlying physical reality without denying that one formalism is more general, or physically deeper, than another. In physics, such theoretical differences can be traced to physical assumptions which approximate certain processes to clarify other processes. I will argue that several approaches to visual perception make approximations which do not accurately represent the physical processes which they have set out to explain. Due to this fact, these theories experience predictive limitations which do not permit them to understand, even in first approximation, major properties of the data. In other words, the mathematical formalisms of these theories have not incorporated fundamental physical intuitions into their computational structure. Once these intuitions are translated into a suitable formalism, the theoretical diversity in visual science will, I claim, gradually become qualitatively more like that known in physics.

The comparison with physics is not an idle one. Certain of the intuitions which need to be formalized at the foundations of visual theory are well known to us all. They have not been acted upon because, despite their simplicity, they lead to conceptually radical conclusions that force a break with traditional notions of geometry. Lines and edges can no longer be thought of as series of points; planes can no longer be built up from local surface elements or from sets

of lines or points; and so on. All local entities evaporate as we build up notions of functional perceptual units which can naturally deal with the global context-dependent nature of visual percepts. The formalism in which this is achieved is a quantized dynamic geometry, and the nature of the quantization helps to explain why so many visual percepts seem to occur in a curved visual space.

When a physicist discusses quantization of curved space, he usually means joining quantum mechanics to general relativity. This goal has not yet been achieved in physics. To admit that even the simplest visual phenomena suggest such a formal step clarifies both the fragmentation of visual science into physically inadequate formalisms, and the radical nature of the conceptual leap that is needed to remedy this situation.

3. The Need for Theories which Match the Data's Coherence

As background for my theoretical treatment, I will review various paradoxical data concerning interactions between the perceived depth, lightness, and form of objects in a scene. These paradoxes should not, I believe, be viewed as isolated and unimportant anomalies, but rather as informative instances of how the visual system completes a scene's global representation in response to locally ambiguous visual data. These data serve to remind us of the interdependence and context-sensitivity of visual properties; in other words, of their coherence. With these reminders fresh in our minds, I will argue in Part II that by probing important visual design principles on a deep mathematical level, one can discover, as automatic mathematical consequences, how many visual properties are coherently caused as manifestations of these design principles.

This approach to theory construction is not in the mainstream of psychological thinking today. Instead, one often finds models capable of computing some single visual property, such as edges or cross-correlations. Even with a different model for each property, this approach does not suggest how related visual properties work

together to generate a global visual representation. For example, the present penchant for modelling lateral inhibition by linear feedforward operators like a Laplacian or a Fourier transform to compute edges or cross-correlations (Marr and Hildreth, 1980 ; Robson, 1975) pays the price of omitting related nonlinear properties like reflectance processing, Weber law modulation, figure-ground filling-in, and hysteresis. To the argument that one must first understand one property at a time, I make this reply: The feedforward linear theories contain errors even in the analysis of the concepts they set out to explain. Internal problems of these theories prevent them from understanding the other phenomena that cohere in the data.

This lack of coherence, let alone correctness, will cause a heavy price to be paid in the long run, both scientifically and technologically. Unless the relationships among visual data properties are correctly represented in a distributed fashion within the system, plausible (and economic) ways to map these properties into other subsystems, whether linguistic, motor, or motivational, will be much harder to understand. Long-range progress, whether in theoretical visual science per se, or in its relationships to other scientific and technological disciplines, requires that the mathematical formalisms within which visual concepts are articulated be scrupulously criticized.

4. Some Influences of Perceived Depth on Perceived Size

Interactions between an object's perceived depth, size, and lightness have been intensively studied for many years. The excellent texts by Cornsweet (1970) and Kaufman (1974) review many of the basic phenomena.

The classical experiments of Holway and Boring (1941) show that observers can estimate the actual sizes of objects at different distances even if all the objects subtend the same visual angle on the observers' retinas. Binocular cues contribute to the invariant percept of size. For example, Emmert (1881) showed that monocular cues may be insufficient to estimate an object's length. He noted that a mo-

nocular afterimage seems to be located on any surface which the subject binocularly fix-
-ates while the afterimage is active. Moreover, the perceived size of the after-
image increases as the perceived distance of the surface increases. This effect
is called Emmert's Law.

Gogel (1956, 1965, 1970) has reported that two objects viewed under reduc-
tion conditions (one eye looks through a small aperture in dim light) will be more
likely to be judged as equidistant from the observer as they are brought closer
together in the frontal plane. In a related experiment, one object is monocularly
viewed through a mirror arrangement, whereas all other objects in the scene are
binocularly viewed. The monocularly viewed object then seems to lie at the same
distance as the edge among all the binocularly viewed objects that is retinally most
contiguous to it. Gogel interpreted these effects as examples of an equidistance
tendency in depth perception. The equidistance tendency also holds if a monocular
afterimage occupies a retinal position near to that excited by a binocularly viewed
object. The perceived distance of the binocular object influences the perceived
distance of the adjacent afterimage by the equidistance tendency, and thereupon in-
fluences the perceived size of the afterimage by Emmert's Law.

Results such as these show that depth cues exert a powerful influence on size
estimates. These results also suggest that this influence can propagate between ob-
ject representations whose cues excite disparate retinal points and that the pat-
terning of all cues in the visual context of an object helps to determine its per-
ceived length. The classical geometric notion that length can be measured by a ruler,
or can be conceptualized in terms of any locally defined computation, hereby falls
into jeopardy.

5. Some Monocular Constraints on Size Perception

Size estimates can also be modified by monocular cues, as in the corridor illusion (Richards and Miller, 1971). In this illusion, two cylinders of equal size in a picture are perceived to be of different sizes because they lie in distinct positions within a rectangular grid whose spatial scale diminishes towards a fixation point on the horizon. An analogous effect occurs in the Ponzo illusion, wherein two horizontal rods of equal pictorial length are drawn superimposed over an inverted V (Kaufman, 1974). The upper rod appears longer than the lower rod. The perception of these particular figures may be influenced by learned perspective cues to depth (Gregory, 1966), although this hypothesis does not explain how perspective cues alter length percepts. There exist many other figures, however, wherein a perspective effect on size scaling is harder to rationalize (Day, 1972). Several authors have therefore modelled these effects in terms of intrinsic scaling properties of the visual metric (Dodwell, 1975; Eijkman, Jongsma, and Vincent, 1981; Restle, 1971; Watson, 1978).

A more dramatic version of scaling is evident when subjective contours complete the boundary of an incompletely represented figure. Then objects of equal pictorial size that lie inside and outside the completed figure may appear to be of different size (Coran, 1972). The very existence of subjective contours raises the issue of how incomplete data about form can select internal representations which can span, or fill-in, the incomplete regions of the figure. How can we characterize those features or spatial scales in the incomplete figure which play an informative role in the completion process vs. those features or scales which are irrelevant? Attneave (1954) has shown, for example, that when a drawing of a cat is replaced by a drawing in which the points of maximum curvature in the original drawing are joined by straight lines, then the new drawing still looks like a cat. Why are the points of maximum curvature such good indicators of the entire form? Is there a natural reason why certain spatial scales in a figure might have greater weight than other

scales? Attneave's cat raises the question: Why does interpolation between points
of maximum curvature with lines of zero curvature produce a good facsimile of the
original picture? Somehow different spatial scales need to interact in our original
percept for this to happen. To understand this issue, we need a correct definition
of spatial scale. Such a definition should distinguish between local scaling effects,
such as those which can be understood in terms of a neuron's receptive field (Robson,
1975), and global scaling effects, such as those which control the filling-in of sub-
jective contours or of phantom images across a movie screen which subtends a visual
angle much larger than that spanned by any neuron's receptive field (Smith and Over,
1979; Tynan and Sekuler, 1975; von Grunau, 1979; Weisstein, Maguire, and Berbaum,
1977).

6. Multiple Scales in Figure and Ground: Simultaneous Fusion and Rivalry

That interactions between several spatial scales are needed for form percep-
tion is also illustrated by the following type of demonstration (Beck, 1972). Rep-
resent a letter E by a series of nonintersecting straight lines of varying ob-
lique and horizontal orientations drawn within an imaginary E contour and sur-
rounded by a background of regular vertical lines. The E is not perceived because
of the lines within the contour, because the several orientations of these inter-
ior lines do not group into an E-like shape. Somehow the E is synthesized as the
complement of the regular background, or more precisely by the statistical dif-
ferences between the figure and the ground . These statistical regularities
define a spatial scale — broader than the scale of the individual lines — on
which the E can be perceived.

In a similar vein, construct a stereogram out of two pictures as follows
(Kaufman, 1974). The left picture is constructed from 45°-oblique dark parallel
lines bounded by an imaginary square, which is surrounded by 135°-oblique lighter
parallel lines. The right picture is constructed from 135°-oblique dark parallel

lines bounded by an imaginary square whose position in the picture is shifted relative to the square in the left picture. This imaginary square is surrounded by 45°-oblique lighter parallel lines. When these pictures are viewed through a stereoscope, the dark oblique lines within the square are rivalrous. Nonetheless, the square as a whole is seen in depth. How does this stereogram induce rivalry on the level of the narrowly tuned scales that interact preferentially with the lines, yet simultaneously generate a coherent depth impression on the broader spatial scales that interact preferentially with the squares?

Kulikowski (1978) has also studied this phenomenon by constructing two pairs of pictures which differ in their spatial frequencies. Each picture is bounded by the same frame, as well as by a pair of short vertical reference lines attached to the outside of each frame at the same spatial locations. In one pair of pictures, spatially blurred black and white vertical bars of a fixed spatial frequency are 180° out of phase. In the other pair of pictures, sharp black and white vertical bars of the same spatial extent are also 180° out of phase. The latter pair of pictures contains high spatial frequency components (edges) as well as low spatial frequency components. During binocular viewing, the subjects can fuse the two spatially blurred pictures and see them in depth with respect to the fused images of the two frames. By contrast, the subjects experience binocular rivalry when they view the two pictures of sharply etched bars. Yet they still experience the rivalrous patterns in depth. This demonstration suggests that the low spatial frequencies in the bar patterns can be fused to yield a depth impression even while the higher spatial frequency components in the bars elicit an alternating rivalrous perception of the monocular patterns.

The demonstrations of Kaufman (1974) and Kulikowski (1978) raise many interesting questions. The most pressing question is perhaps: Why are fusion and ri-

valry alternative binocular perceptual modes? Why are coexisting unfused mono-
cular images so easily supplanted by rivalrous monocular images? How does fusion
at one spatial scale coexist with rivalry at a different spatial scale that re-
presents the same region of visual space?

7. Binocular Matching, Competitive Feedback, and Monocular Self-Matching

These facts suggest some conclusions that will be helpful to organize my data
review and that will be derived on a different theoretical basis in Part II. I will
indicate how rivalry suggests the existence of binocular cells that can be activated
by a single monocular input and that mutually interact in a competitive feedback
network. First I will indicate why these binocular cells can be monocularly activa-
ted.

The binocular cells in question are the spatial loci where monocular data
from the two eyes interact to generate fusion or rivalry as the outcome. To show
why at least some of these cells can be monocularly activated, I will consider
implications of the following mutually exclusive possibilities: either the outcome
of binocular matching feeds back towards the monocular cells that generated the
signals to the binocular cells, or it does not.

Suppose not. Then the activities of monocular cells cannot subserve percep-
tion; rather, perception is associated with activities of binocular cells or of
cells more central than the binocular cells. This is because both sets of mono-
cular cells would remain active during a rivalry percept, since the binocular
interaction leading to the rivalry percept does not, by hypothesis, feed back
to alter the activities of the monocular cells. Now we confront the conclusion
that monocular cells do not subserve perception with the fact that the visual
world can be vividly seen through a single eye. It follows that some of the
binocular cells which subserve perception can be activated by input from a
single eye.

Having entertained the hypothesis that the outcome of binocular matching does
not feed back towards monocular cells, let us now consider the opposite hypothesis.

In this case, too, I will show that a single monocular representation must be able to activate certain binocular cells. To demonstrate this fact, I will again argue by contradiction.

Suppose not. In other words, suppose that the outcome of binocular matching does feed back towards monocular cells but a single monocular input cannot activate binocular cells. Because the visual world can be seen through a single eye, it follows that the activities of monocular cells subserve perception in this case. Consequently, during a binocular rivalry percept, the binocular-to-monocular feedback must quickly inhibit one of the monocular representations. The signals which this monocular representation was sending to the binocular cells are thereupon also inhibited. The binocular cells then receive signals only from the other monocular representation. The hypothesis that binocular cells cannot fire in response to signals from only one monocular representation implies that the binocular cells shut off, along with all of their output signals. The suppressed monocular cells are then released from inhibition and are excited again by their monocular inputs. The cycle can now repeat itself, leading to the percept of a very fast flicker of one monocular view superimposed upon the steady percept of the other monocular view. This phenomenon does not occur during normal binocular vision. Consequently, the hypothesis that a single monocular input cannot activate binocular cells must be erroneous. Whether or not the results of binocular matching feed back towards monocular cells, certain binocular cells can be activated by a single monocular representation.

An additional conclusion can be drawn in the case wherein the results of binocular matching can feed back towards monocular cells. In this case, a single monocular source can activate binocular cells, which can thereupon send signals towards the monocular source. The monocular representation can hereby self-match at the monocular source using the binocular feedback as a matching signal. This fact implies that the monocular source cells are themselves binocular cells, because a monocular input can activate binocular cells which then send feedback signals to the monocular source cells of the other eye. In this way the monocular source cells can be activa-

ted by both eyes, albeit less symmetrically than the binocular cells at which the primary binocular matching event takes place.

This conclusion can be summarized as follows: The binocular cells at which binocular matching takes place are flanked by binocular cells that satisfy the following properties: (a) they are fed by monocular signals; (b) they excite the binocular matching cells; (c) they can be excited or inhibited due to feedback from the binocular matching cells, depending upon whether fusion or rivalry occur.

It remains only to consider the possibility that the results of binocular matching do not feed back towards the monocular cells. The following argument indicates why this cannot happen. A purely feedforward interaction from monocular towards binocular cells cannot generate the main properties of rivalry, namely, a sustained monocular percept followed by rapid and complete suppression of this percept when it is supplanted by the other monocular percept. This is because the very activity of the perceived representation must be the cause of its habituation and loss of competitive advantage relative to the suppressed representation. Consequently, the habituating signals from the perceived representation that inhibit the suppressed representation reach the latter representation at a stage at, or prior to, this representation's locus for generating signals to the other representation that are capable of habituating. Such an arrangement allows the signals of the perceived representation to habituate but spares the suppressed representation from habituation. By symmetry, the two representations reciprocally send signals to each other that are received at, or at a stage prior to, their own signalling cells. This arrangement of signalling pathways defines a feedback network.

One can now refine this conclusion by going through arguments like those above to conclude that (a) the feedback signals are received at binocular cells rather than at monocular cells, and (b) the feedback signals are not all inhibitory signals or else binocular fusion could not occur. Thus a competitive balance between excitatory and inhibitory feedback signals among binocular cells capable of monocular activation needs to be considered. Given the possibility of monocular self-matching

in this framework, one also needs to ask why the process of monocular self-matching, in the absence of a competing input from the other eye, does not cause the cyclic strengthening and weakening of monocular activity that occurs when two nonfused monocular inputs are rivalrous?

One does not need a complete theory of these properties to conclude that no theory in which only a feedforward flow of visual patterns from monocular to binocular cells occurs, say to compute disparity information, can explain these data. Feedback from binocular matching towards monocular computations is needed to explain rivalry data, just as such feedback is needed to explain the influence of perceived depth on perceived size or brightness. I will suggest in Part II how a suitably defined feedback scheme can cause all of these phenomena at once.

8. Against the Keplerian View: Scale-Sensitive Fusion and Rivalry

The Kaufman (1974) and Kulikowski (1978) experiments also argue against the Keplerian view, which is a mainstay of modern theories of stereopsis. The Keplerian view is a realist hypothesis which suggests that the two monocular views are projected point-by-point along diagonal rays, and that their crossing-points are loci from which the real depth of objects may be computed (Kaufman, 1974). When the imaginary rays of Kepler are translated into network hardware, one is led to assume that network pathways carrying monocular visual signals merge along diagonal routes (Sperling, 1970). The Keplerian view provides an elegant way to think about depth, because objects which are closer should, other things being equal, have larger disparities, and their Keplerian pathways should therefore cross at points which are further along the pathways. Moreover all pairs of points with the same disparity cross at the same distance along their pathway, and hereby form a row of contiguous crossing-points.

This concept does not explain a result such as Kulikowski's, since all points in each figure (so the usual reasoning goes) have the same disparity with respect to the corresponding point in the other figure. Hence all points cross in the same

row. In the traditional theories, this means that all points should match equally well to produce an unambiguous disparity measure. Why, then, do low spatial frequencies seem to match and yield a depth percept at the same disparity at which high spatial frequencies do not seem to match?

Rather than embrace the Keplerian view, I will suggest how suitably proprocessed input data of fixed disparity can be matched by certain spatial scales but not by other spatial scales. To avoid misunderstanding, I should immediately say what this hypothesis does not imply. It does not imply that a pair of high spatial frequency input patterns of large disparity cannot be matched, because only suitable statistics of the monocular input patterns will be matched, rather than the input patterns themselves. Furthermore, inferences made from linear statistics of the input patterns do not apply because the statistics in the theory need to be nonlinear averages of the input patterns to ensure basic stability properties of the feedback exchange between monocular and binocular cells. These assertions will be clarified in Part II.

Once the Keplerian view is questioned, then the problem of false-images (Julesz, 1971) which derives from this view, and which has motivated much thinking about stereopsis, also becomes less significant. The false-images are those crossing-points in Kepler's grid that do not correspond to the objects' real disparities.

Workers like Marr and Poggio (1979) have also concluded that false images are not a serious problem if spatial scaling is taken into account. Their definition of spatial scale differs from my own in a way that highlights how a single formal definition can alter the whole character of a theory. For example, when they mixed their definition of a spatial scale with their view of the false-image problem, Marr and Poggio (1979) were led to renounce cooperativity as well, which I view as an instance of throwing out the baby with the bathwater, since all global filling-in and figure-ground effects hereby become inexplicable in their theory. Marr and Poggio (1979) abandoned cooperativity because they did not need it to deal with false images. In a model such as theirs whose primary goal is to compute unambiguous disparity measures, their conclusion seems quite logical. Confronted by the greater body of phenomena

that are affected by depth estimates, such a step seems unwarranted.

9. Local vs. Global Spatial Scales

Indeed, both the Kaufman (1974) and the Kulikowski (1978) experiments, among many others, illustrate that a figure or ground has a coherent visual existence that is more than the sum of its unambiguous feature computations. Once a given spatial scale makes a good match in these experiments, a depth percept is generated that pervades a whole region. We therefore need to distinguish between the scaling property that makes good matches based on local computations from the global scaling effects that fill-in an entire region subtending an area much broader than the local scales themselves.

This distinction between local and global scaling effects is vividly demonstrated by constructing a stereogram in which the left "figure" and its "ground" are both induced by a 5% density of random dots (Julesz, 1971, p. 336), and the right "figure" of dots is shifted relative to its position in the left picture. Stereoscopically viewed, the whole figure, including the entire 95% of white background between its dots, seems to hover at the same depth. How does the white background of the "figure" inherit the depthfulness due to the disparities of its meagerly distributed dots, and the white background of the "ground" inherits the depthfulness of its dots? What mechanism organizes the locally ambiguous white patches that dominate 95% of the pictorial area into two distinct and internally coherent regions? Julesz (1971, p. 256) describes another variant of the same phenomenon using a random-dot stereogram inspired by an experiment of Shipley (1965). In this stereogram, the traditional center square in depth is interrupted by a horizontal white strip that cuts both the center square and the surround in half. During binocular viewing, the white strip appears to be cut along the contours of the square and it inherits the depth of figure or ground, despite the fact that it provides no disparity or brightness cues of its own at the cut regions.

10. Interaction of Perceived Form and Perceived Position

The choice of scales leading to a depth percept can also cause a shift in perceived form, notably in the relative distance between patterns in a configuration. For example, when a pattern AB C is viewed through one eye and a pattern A BC is viewed through the other eye, the letter B can be seen in depth at a position halfway between A and C (Von Tschermak-Seysenegg, 1952; Werner, 1937). This phenomenon, called displacement or allelotropia, again suggests that the dynamic transformations within visual space are not of a local character since the location of entire letters, let alone their points and lines, can be deformed by the spatial context in which they are placed. The non-local nature of visual space extends also to brightness perception, as the next section reviews.

11. Some Influences of Perceived Depth and Form on Perceived Brightness

The Craik-O'Brien and Cornsweet effects (Cornsweet, 1970; O'Brien, 1958) show that an object's form, notably its edges or regions of rapid spatial change, can influence its apparent brightness, or lightness (Figure 1). Let the luminance

Figure 1

profile in Figure 1a describe a cross-section of the 2-dimensional picture in Figure 1b. Then the lightness of this picture appears as in Figure 1c. The edges of the luminance profile determine the lightnesses of the adjacent regions by a filling-in process. Although the luminances of the regions are the same except near their edges, the perceived lightnesses of the regions are determined by the brightnesses of their respective edges. This remarkable property is reminiscent of Attneave's cat, since regions of maximum curvature — in the lightness domain — again help to determine how the percept is completed. In the present instance, the filling-in process overrides the visual data rather than merely completing an incomplete pattern.

Hamada (1976, 1980) has shown that this filling-in process is even more para-doxical than was previously thought. He compared the lightness of a uniform background with the lightness of the same uniform background with a less luminous Craik-O'Brien figure superimposed on it. By the usual rules of brightness contrast, the lesser brightness of the Craik-O'Brien figure should raise the lightness of the background as its own lightness is reduced. Remarkably, even the background seems darker than the uniform background of the comparison figure, although its luminance is the same.

Just as form can influence lightness, so too can apparent depth influence lightness. Figures which appear to lie at the same depth can influence each other's lightness in a manner analogous to that found in a monocular brightness constancy paradigm (Gilchrist, 1979).

12. Some Influences of Perceived Brightness on Perceived Depth

Just as depth can influence brightness estimates, brightness data can influence depth estimates. For example, Kaufman, Bacon, and Barroso (1973) studied stereograms built up from the two monocular pictures in Figure 2a. When these pictures are viewed through a stereogram, the eyes see the lines at a different depth due to

Figure 2

the disparity between the two monocular views. If the stereogram is changed so that the left eye sees the same picture as before, whereas the right eye sees the two pictures superimposed (Figure 2b), then depth is still perceived. If both eyes see the same superimposed pictures, then of course no depth is seen. However, if one eye sees the pictures superimposed with equal brightness, whereas the other eye sees the two pictures superimposed, one with less brightness and

the other with more brightness, then depth is again seen. In this latter case, there
is no disparity between the two figures, although there is a brightness difference.
How does this brightness difference elicit a percept of depth?

The Kaufman et al (1973) study raises an interesting possibility. If a binocu-
lar brightness difference can cause a depth percept and if a depth percept can influ-
ence perceived length, then a binocular brightness difference should be able to cause
a change in perceived length. It is also known that monocular cues can sometimes have
similar effects on perceived length as binocular cues, as in the corridor and Ponzo
illusions. When these two phenomena are combined, it is natural to ask: Under what
circumstances can a monocular brightness change cause a change, albeit small, in per-
ceived length? I will return to this question in Part II.

13. The Binocular Mixing of Monocular Brightnesses

The Kaufman et al (1973) result illustrates the fact that brightness informa-
tion from each eye somehow interacts in a binocular exchange. That this exchange is
not simply additive is shown by several experiments. For example, let AB on a white
field be viewed with the left eye and BC on a white field be viewed with the right
eye in such a way that the two B's are superimposed. Then the B does not look sig-
nificantly darker than A and C despite the fact that white is the input to the other
eye corresponding to these letter positions (Helmholtz, 1962). In a similar vein,
closing one eye does not make the world look half as bright despite the fact that
the total luminance reaching the two eyes is halved (Levelt, 1964; Von Tschermak-
Seysenegg, 1952). This fact recalls the discussion of monocular firing of binocular
cells from Section 7.

The subtlety of binocular brightness interactions is further revealed by Fech-
ner's Paradox (Hering, 1964). Suppose that a scene is viewed through both eyes, but that
one eye sees the scene through a neutral filter that attenuates all wavelengths by a
constant ratio. The filter does not distort the reflectances, or ratios, of light
reaching its eye, but only its absolute intensity. Now let the filtered eye be entire-

ly occluded. Then the scene looks brighter and more vivid despite the fact that less total light is reaching the two eyes, and the reflectances are still the same.

Binocular summation of brightness, in excess of probability summation, can occur when the monocular inputs are suitably matched "within some range, perhaps equivalent to Panum's area...stereopsis and summation may be mediated by a common neural mechanism" (Blake, Sloane, and Fox, 1981). I will suggest below that the co-existence of Fechner's paradox and binocular brightness summation can be explained by properties of binocular feedback exchanges among multiple spatial scales. This explanation provides a theoretical framework in which recent studies and models of interactions between binocular brightness summation and monocular flashes can be interpreted (Cogan, Silverman, and Sekuler, 1982).

Wallach and Adams (1954) have shown that if two figures differ only in terms of the reflectance of one region, then quite the opposite of summation may be found. In fact, a rivalrous perception of brightness can be generated in which one shade, then the other, is perceived rather than a simultaneous average of the two shades. I will suggest below that this rivalry phenomenon may be related to the possibility that two monocular figures of different lightness may generate different spatial scales and thereby create a binocular mismatch.

Having reviewed some data concerning the mutual interdependence and lability of depth, form, and lightness judgements, I will now review some obvious visual facts that seem paradoxical when placed beside some of the theoretical ideas that are in vogue at this time. I will also point out that some popular and useful theoretical approaches are inherently limited in their ability to explain either these paradoxes or the visual interactions summarized above.

14. The Insufficiency of Disparity Computations

It is a truism that the retinal images of objects at optical infinity have zero disparity, and that as an object approaches an observer, the disparities on the two retinas of corresponding object points tend to increase. This is the commonplace rea-

son for assuming that larger disparities are an indicator of relative closeness.
Julesz stereograms (Julesz, 1971) have, moreover, provided an elegant paradigm
wherein disparity computations are a sufficient indicator of depth, since each
separate Julesz random dot picture contains no monocular form cues, yet statisti-
cally reliable disparities between corresponding random dot regions yield a vivid
impression of a form hovering in depth.

This stunning demonstration has encouraged a decade of ingenious neural model-
ing. Sperling (1970) introduced important pioneering concepts and equations in a
classic paper that explains how cooperation within a disparity plane and competition
between disparity planes can resolve binocular ambiguities. These ideas were deve-
loped into an effective computational procedure in Dev (1975) which led to a number
of mathematical and computer studies (Amari and Arbib, 1977; Marr and Poggio, 1976).
Due to these historical considerations, I will henceforth call models of this type
Sperling-Dev models.

All Sperling-Dev models assume that corresponding to each small retinal
region there exist a series of disparity detectors sensitive to distinct dispari-
ties. These disparity detectors are organized in sheets such that cooperative
effects occur between detectors of like disparity within a sheet, whereas compe-
titive interactions occur between sheets. The net effect of these interactions is
to suppress spurious disparity correlations and to carve out connected regions of
active disparity detectors within a given sheet. These active disparity regions
are assumed to correspond to a depth plane of the underlying retinal regions. Some
investigators have recently expressed their enthusiasm for this interpretation by
committing the homuncular fallacy of drawing the depth planes in impressive 3-di-
mensional figures which carry the full richness of the monocular patterns, although
within the model the monocular patterns do not differentially parse themselves among
the several sheets of uniformly active disparity detectors.

That something is missing from these models is indicated by the following con-
siderations. The use of a stereogram composed from two separate pictures does not

-20-

always well approximate the way two eyes view a single picture. When both eyes

focus on a single point within a patterned planar surface viewed in depth, the

fixation point is a point of minimal binocular disparity. Points increasingly

far from the fixation point have increasingly large disparities. Why doesn't

such a plane recede towards optical infinity at the fixation point, and curve

towards the observer at the periphery of the visual field? Why doesn't the plane

get distorted in a new way every time our eyes fixate on a different point within

its surface? If disparities are a sufficient indicator of depth, then how do we

ever see planar surfaces? Or even rigid surfaces?

This insufficiency cannot be escaped just by saying that an observer's spatial

scales get bigger as retinal eccentricity increases. To see this, let a bounded

planar surface have an interior which is statistically uniform with respect to an

observer's spatial scales (in a sense that will be precisely defined in Part II).

Then the interior disparities of the surface are ambiguous. Only its boundary dis-

parities supply information about the position of the surface in space. Filling-in

between these boundaries to create a planar impression is not just a matter of

showing that the same disparity, even after an eccentricity compensation, can be

locally computed at all the interior points, because an unambiguous disparity compu-

tation cannot be carried out at the interior points. The issue is not just whether

the observer can estimate the depth of the planar surface, but also how the obser-

ver knows that a planar surface is being viewed.

This problem is hinted at even when Julesz stereograms are viewed. Staring at

one point in the stereogram results in the gradual loss of depth (Kaufman, 1974).

Also in a stereogram composed of three vertical lines to the left eye and just the

two flanking lines to the right eye, the direction of depth of the middle line depends

on whether the left line or the right line is fixated (Kaufman, 1974). This demonstra-

tion makes the problem of perceiving planes more severe for any theory which restricts

itself to disparity computations, since it shows that depth can depend on the fixation

points. What is the crucial difference between the way we perceive the depths of lines

and planes? Kaufman (1974) seems to have had this problem in mind when he wrote "all theories of stereopsis are really inconsistent with the geometry of stereopsis" (p. 320).

Another problem faced by Sperling-Dev models is that they cannot explain effects of perceived depth on perceived size and lightness. The attractive property that the correct depth plane fills-in with uniform activity due to local cooperativity creates a new problem: How does the uniform pattern of activity within a disparity plane rejoin the nonuniformly patterned monocular data to influence its apparent size and lightness?

Finally there is the problem that there can only exist a finite number of depth planes in a finite neural network. Only a few such depth planes can be inferred to exist by joining data relating spatial scales to perceived depth, such as the Kaufman (1974) and Kulikowski (1978) data summarized in Section 6, to spatial frequency data which suggest that only a few spatial scales exist (Graham, 1981; Wilson and Bergen, 1979). Since only one depth plane is allowed to be active at each time in any spatial position in a Sperling-Dev model, apparent depth should discretely jump a few times as an observer approaches an object. Instead, apparent depth seems to change continuously in this situation.

15. The Insufficiency of Fourier Models

An approach with a strong kernel of truth but a fundamental predictive limitation is the Fourier approach to spatial vision. The kernel of truth is illustrated by threshold experiments with four different types of visual patterns (Graham and Nachmias, 1971; Graham, 1981). Two of the patterns are gratings which vary sinusoidally across the horizontal visual field with different spatial frequencies. The other two patterns are the sum and difference patterns of the first two patterns. If the visual system behaved like a single channel wherein larger peak-to-trough pattern intensities are more detectable, the compound patterns would be more detectable than the sinusoidal patterns. In fact, all the patterns are approximately equally detectable.

A model wherein the different sinusoidal spatial frequencies are independently filtered by separate spatial channels, or scales, fits the data much better. Recall from Section 6 some of the other data that also suggests the existence of multiple scales.

A related advantage of the multiple channel idea is that one can filter a complex pattern into its component spatial frequencies, weight each component with a factor that mirrors the sensitivity of the human observer to that channel, and then resynthesize the weighted pattern and compare it with an observer's perceptions. This modulation transfer function approach has been used to study various effects of boundary edges on interior lightnesses (Cornsweet, 1970). If the two luminance profiles in Figure 3 are filtered in this way, then they both generate the same output pattern

Figure 3

because the human visual system attenuates low spatial frequencies. Unfortunately both output patterns look like a Cornsweet profile, whereas actually the Cornsweet profile looks like a rectangle. This is not a minor point, since the interior regions of the Cornsweet profile have the same luminance, which is false in the rectangular figure.

This application of the Fourier approach seems to be a misplaced one to me, since the Fourier transform is a linear transformation, whereas a reflectance computation must involve some sort of ratios and is therefore inherently nonlinear.

The Fourier scheme is also a feedforward transformation of an input pattern into an output pattern. It cannot, in principle, explain how apparent depth alters apparent length and brightness, since such computations depend on a feedback exchange between monocular data to engender binocular responses. In particular, the data reviewed in Section 4 show that the very definition of a length scale can remain ambiguous until it is embedded in a binocular feedback scheme. The Fourier transform does not at all suggest why length estimates should be so labile. The multiple channel and sensitivity notions need to be explicated in a different formal framework.

16. The Insufficiency of Linear Feedforward Theories

The above criticisms of the Fourier approach to spatial vision hold for all computational theories that are based on linear and feedforward operations. For example, some recent workers in artificial intelligence (Marr and Hildreth, 1980), compute a spatial scale by first linearly smoothing a pattern with respect to a Gaussian distribution, and compute an edge by setting the Laplacian (the second derivatives) of the smoothed pattern equal to zero (Figure 4). The use of the La-

Figure 4

placian to study edges goes back at least to the time of Mach (Ratliff, 1965). The Laplacian is time-honored, but it suffers from limitations that become more severe when its zero-crossings are made the center-piece of a theory of edges.

One of many difficulties is that a zero-crossing computation computes only the position of an edge, and not other related properties such as the brightness of the pattern near the edge. However, the Cornsweet and Craik-O'Brien figures pointedly show that the brightnesses of edges can strongly influence the lightness of their enclosed forms. Something more than zero-crossings is therefore needed to understand spatial vision. The zero-crossing computation itself does not disclose what is missing, so its advocates must guess what is needed. Marr and Hildreth (1980) guess that factors like position, orientation, contrast, length, and width should be computed at the zero-crossings. These guesses do not follow from their definition of an edge, or from their computation of an edge. Such properties lie beyond the implications of the zero-crossing computation, because this computation discards essential features of the pattern near the zero-crossing location. Even if the other properties are added on to a list of data that is stored in computer memory, this list distorts, indeed entirely destroys, the intrinsic geometric structure of the pattern. The replacement of the natural internal geometrical relationships of a pattern by arbitrary numerical measures of the pattern prevents the Marr and Hildreth (1980) theory from understanding how global processes, such as filling-in, can spontaneously occur in a physical

setting. Instead, the Marr and Hildreth (1980) formulation leads to an approach
wherein all the intelligence of what to do next rests in the investigator rather
than in the model. This restriction to local investigator-driven computations is
due not only to the present state of their model's development, but also to the
philosophy of these workers, since Marr and Hildreth (1980) write (p. 189): "the
visual world is not constructed of ripply, wave-like primitives that extend and
add together over an area." Finally, because their theory is linear, it cannot tell
us how to estimate the lightnesses of objects, and because their theory is feedfor-
ward, it cannot say how apparent depth can influence the apparent size and lightness
of monocular patterns.

17. The Filling-In Dilemma: To Have Your Edge and Fill-In Too

Any linear and feedforward approach to spatial vision is, in fact, confronted
by the full force of the filling-in dilemma: If spatial vision operates by first
attenuating all but the edges in a pattern, then how do we ever arrive at a percept
of rigid bodies with ample interiors, which are, after all, the primary objects of
perception? How can we have our edges and fill-in too? How does the filling-in pro-
cess span retinal areas which far exceed the spatial bandwidths of the individual
receptive fields that physically justify a Gaussian smoothing process? In particular,
in the idealized luminance profile in Figure 5, after the edges are determined by a

Figure 5

zero-crossing computation, the directions in which to fill-in are completely ambigu-
ous without further computations tacked on. I will argue in Part II of the article
that a proper definition of edges does not require auxiliary guesswork.

I should emphasize what I do not mean by a solution to the filling-in dilemma.
It is not sufficient to say that edge outlines of objects constitute sufficient infor-
mation for a viewer to understand a 3-dimensional scene. Such a position merely says
that observers can use edges to arrive at object percepts, but not how they do so.

such a view begs the question. It is also not sufficient to say that feedback
expectancies, or hypotheses, can use edge information to complete an object per-
cept. Such a view does not say how the feedback expectancies were learned, notably
what substrate of completed form information was sampled by the learning process.
This view also begs the question. Finally, it is inadequate to say that an abstract
reconstruction process generates object representations from edges if this process
would require a homunculus for its execution in real-time.

Expressed in another way, the filling-in dilemma asks: If it is really so hard
for us to find mechanisms which can spontaneously and unambiguously fill-in between
edges, then do we not have an imperfect understanding of why the nervous system
bothers to compute edges? Richards and Marr (1981) suggest that the edge computation
compresses the amount of data which needs to be stored. This sort of memory load
reduction is important in a computer program, but I will suggest in Part II that it
is not a rate-limiting constraint on the brain design which grapples with binocular
data. I will suggest, by contrast, that the edge computation sets the stage for pro-
cesses which selectively amplify and fill-in among those aspects of the data which
are capable of matching monocularly, binocularly and/or with learned feedback expec-
tancies, as the case might be.

PART II

18. Edges and Fixations: The Ambiguity of Statistically Uniform Regions

I will motivate my theoretical constructions with two simple gedanken experiments. I will use these experiments to quickly remind us of some important relationships between perceived depth and the monocular computation of spatial nonuniformities.

Suppose that an observer attempts to fixate a perceptually uniform rectangle hovering in space in front of a discriminable but perceptually uniform background. How does the observer know where to fixate the rectangle? Even if each of the observer's eyes independently fixates a different point of the rectangle's interior, both eyes will receive identical input patterns near their fixation points due to the rectangle's uniformity. The monocular visual patterns near the fixation points match no matter how disparately the fixation points are chosen within the rectangle.

Several conclusions follow from this simple observation. Binocular visual matching between spatially homogeneous regions contains no information about where the eyes are printed, since all binocular matches between homogeneous regions are equally good no matter where the eyes are pointed. The only binocular visual matches which stand out above the baseline of ambiguous homogeneous matches across the visual field are those which correlate spatially nonuniform data to the two eyes. However, the binocular correlations between these nonuniform patterns, notably their disparities, depend upon the fixation points of the two eyes. Disparity information by itself is therefore insufficient to determine the object's depth. Instead, there must exist an interaction between vergeance angle and disparity information to determine where an object is in space (Foley, 1980; Grossberg, 1976; Marr and Poggio, 1979; Sperling, 1970).

This binocular constraint on resolving the ambiguity of where the two eyes are looking is one reason for the monocular extraction of the edges of a visual form

and the attendant suppression of regions which are spatially homogeneous with respect to a given spatial scale. Without the ability to know where the object is in space, there would be little evolutionary advantage in perceiving its solidity or interior. In this limited sense, edge detection is more fundamental than form detection in dealing with the visual environment.

Just knowing that a feedback loop must exist between motor vergeance and sensory disparities does not determine the properties of this loop. Sperling (1970) has postulated that vergeance acts to minimize a global disparity measure. Such a process would tend to reduce the perception of double images (Kaufman, 1974). I have suggested (Grossberg, 1976) that good binocular matches generate an amplification of network activity, or a binocular resonance. An imbalance in the total resonant output from each binocular hemifield may be an effective ver-geance signal leading to hemifield-symmetric resonant activity which signifies good binocular matching and stabilizes the vergeance angle. The theoretical sec-tions below will suggest how these binocular resonances also compute coherent depth, form, and lightness information.

19. Object Permanence and Multiple Spatial Scales

The second gedanken experiment reviews a use for multiple spatial scales, rather than a single edge computation, corresponding to each retinal point. Again our conclusions can be phrased in terms of the fixation process.

As a rigid object approaches an observer, the binocular disparities between its nonfixated features increase proportionally. In order to achieve a concept of object permanence, and at the very least to maintain the fixation process, mechan-isms capable of maintaining a high correlation between these progressively larger disparities are needed. The largest disparities will, other things being equal, lie at the most peripheral points on the retina. The expansion of spatial scales with retinal eccentricity is easily rationalized in this way (Hubel and Wiesel, 1977; Richards, 1975, Schwartz, 1980).

It does not suffice, however, to posit that a single scale exists at each retinal position such that scale size increases with retinal eccentricity. This is because objects of different size can approach the observer. As in the Holway and Boring (1941) experiments, objects of different size can generate the same retinal image if they lie at different distances. If these objects possess spatially uniform interiors, then the boundary disparities of their monocular retinal images carry information about their depth. Because all the objects are at different depths, these distinct disparities need to be computed with respect to that retinal position in one eye that is excited by all the objects' boundaries. Multiple spatial scales corresponding to each retinal position can carry out these multiple disparity computations. How the particular scales which can binocularly resonate to a given object's monocular boundary data thereupon fill-in the internal homogeneity of the object's representation with length and lightness estimates will now be discussed, along with the related question of how monocular cues and learned expectancies can induce similar resonances and thus a perception of depth.

20. Cooperative vs. Competitive Binocular Interactions

One major difference between my approach to these problems and alternative approaches is the following. I suggest that a competitive process, not a cooperative process, defines a depth plane. The cooperative process that other authors have envisaged leads to sheets of network activity which are either off or maximally on. The competitive process that I posit can sustain quantized patterns of activity that reflect an object's perceived depth, lightness, and length. In other words, the competitive patterns do not succumb to a homuncular dilemma. They are part of the representation of an object's binocular form. The cells that subserve this representative process are sensitive to binocular disparities, but they are not restricted to disparity computations. In this sense, they do not define a "depth plane" at all.

One reason that other investigators have not drawn this conclusion is because

a <u>binary code</u> hypothesis is often explicit, or lurks implicitly, in their theories. The intuition that a depth plane can be perceived seems to imply cooperation because, in a binary world, competition implies an <u>either-or</u> choice, which is manifestly unsuitable, whereas cooperation implies an <u>and</u> conjunction, which is at least tolerable. In actuality, a binary either-or choice does not begin to capture the properties of a competitive network. Mathematical analysis is needed to understand these properties. I should emphasize at this point that cooperation and cooperativity are not the same notion. Both competitive and cooperative networks exhibit cooperativity, in the sense with which this word is casually used.

A large body of mathematical results concerning competitive networks has been discovered during the past decade (Ellias and Grossberg, 1975; Grossberg, 1970, 1972, 1973, 1978a,b,c,d, 1980a,b, 1981; Grossberg and Levine, 1975; Levine and Grossberg, 1976). These results clarify that not all competitive networks enjoy the properties that are needed to build a visual theory. Certain competitive networks whose cells obey the membrane equations of neurophysiology do have desirable properties. Such systems are called <u>shunting</u> networks to describe the multiplicative relationship between membrane voltages and the conductance changes that are caused by network inputs and signals. This multiplicative relationship enables these networks to automatically retune their sensitivity in response to fluctuating background inputs. Such an automatic gain control property subserves reflectance processing, Weber-law modulation, sensitivity shift properties to different backgrounds, as well as other important visual effects. Most other authors have worked with additive networks, which do not possess the automatic gain control properties of shunting networks. Sperling (1970, 1981) and Sperling and Sondhi (1968) are notable among other workers in vision for understanding the need to use shunting dynamics, as opposed to mere equilibrium laws of the form $I(A+J)^{-1}$. However, these authors did not develop the mathematical theory far enough to have at their disposal some formal properties that I will need. A

review of these and other competitive properties is found in Grossberg (1981, Sections 10-27). The sections below build up concepts leading to binocular resonances.

21. Reflectance Processing, Weber Law Modulation and Adaptation Level in Feed-
    forward Shunting Competitive Networks

Shunting competitive networks can be derived as the solution of a processing dilemma that confronts all cellular tissues, the so-called noise-saturation dilemma (Grossberg, 1973, 1980a). This dilemma notes that accurate processing both of low activity and high activity input patterns can be prevented by sensitivity loss due to noise (at the low activity end) and saturation (at the high activity end) of the input spectrum. Shunting competitive networks overcome this problem by enabling the cells to automatically retune their sensitivity as the overall background activity of the input pattern fluctuates through time. This result shows how cells can adapt their sensitivity to input patterns that fluctuate over a dynamical range that is much broader than the output range of the cells.

As I mentioned above, the shunting laws take the form of the familiar membrane equations of neurophysiology in neural examples. Due to the generality of the noise-saturation dilemma, formally similar laws should occur in nonneural cellular tissues. I have illustrated in Grossberg (1978e) that some principles which occur in neural tissues also regulate nonneural developmental processes for similar computational reasons.

The solution of the noise-saturation dilemma that I will review herein describes intercellular tuning mechanisms. Data describing intracellular adaptation have also been reported (Baylor and Hodgkin, 1974; Baylor, Hodgkin, and Lamb, 1974a,b) and have been quantitatively fit by a model wherein visual signals are multiplicatively gated by a slowly accumulating transmitter substance (Carpenter and Grossberg, 1981). The simplest intercellular mechanism describes a competitive feedforward network in which the activity, or potential, $x_i(t)$ of the ith cell

(population) v_i in a field of cells $v_1, v_2, \ldots, v_n$ responds to a <u>spatial pat-</u> <u>tern</u> $I_i(t) = \theta_i I(t)$ of inputs $i=1,2,\ldots,n$. A collection of inputs comprises a spatial pattern if each input has a fixed relative size (or reflectance) θ_i but a possibly variable background intensity $I(t)$, say due to a fluctuating light source. The convention that $\sum_{k=1}^{n} \theta_k = 1$ implies that $I(t)$ is the total input to the field; viz. $I(t) = \sum_{k=1}^{n} I_k(t)$. The simplest law which solves the noise-saturation dilemma describes the net rate $\frac{dx_i}{dt}$ at which sites at v_i are activated and/or inhibited through time. This law takes the form:

$$\frac{dx_i}{dt} = -Ax_i + (B-x_i)I_i - (x_i+C) \sum_{k \neq i} I_k, \tag{1}$$

$i=1,2,\ldots n$ where $B > 0 \geq -C$ and $B \geq x_i(t) \geq -C$ for all times $t \geq 0$. Term $-Ax_i$ describes the spontaneous decay of activity at a constant rate $-A$. Term $(B-x_i)I_i$ describes the activation due to an excitatory input I_i in the ith channel (Figure 6). Term $-(x_i+C) \sum_{k \neq i} I_k$ describes the inhibition of activity by competitive inputs

$\sum_{k \neq i} I_k$ from the input channels other than v_i.

In the absence of inputs (viz., all $I_i = 0$, $i = 1,2,\ldots n$), the potential decays to the equilibrium potential 0 due to the decay term $-Ax_i$. No matter how intense the inputs I_i are chosen, the potential x_i remains between the values B and $-C$ at all times because $(B-x_i)I_i = 0$ if $x_i = B$ and $-(x_i+C) \sum_{k \neq i} I_k = 0$ if

$x_i = -C$. That is why B is called an excitatory saturation point and $-C$ is called an inhibitory saturation point. When $x_i > 0$, the cell v_i is said to be depolarized. When $x_i < 0$, the cell v_i is hyperpolarized. The cell can be hyperpolarized only if $C > 0$ since $x_i(t) \geq -C$ at all times t.

Before noting how system (1) solves the noise saturation dilemma, I should clarify its role in the theory as a whole. System (1) is part of a mathematical

classification theory wherein a sequence of network variations on the noise-saturation theme is analyzed. The classification theory characterizes how changes in network parameters (e.g., decay rates or interaction rules) alter the transformation from input pattern $(I_1, I_2, \ldots I_n)$ to activity pattern $(x_1, x_2, \ldots x_n)$. This approach provides a precise understanding of how to design networks to accomplish specialized processing tasks. The inverse process of inferring which networks can generate prescribed data properties is thereby greatly facilitated. In the present case of system (1), a feedforward flow of inputs to activities occurs wherein a narrow on-center of excitatory input (term $(B-x_i)I_i$) is balanced against a broad off-surround of inhibitory inputs (term $-(x_i+C) \sum_{k \neq i} I_k$). Deviations from these hypotheses will generate network properties that differ from those found in system (1), as I will note in subsequent examples.

To see how system (1) solves the noise-saturation dilemma, let the background input $I(t)$ be held steady for a while. Then the activities in (1) approach equilibrium. These equilibrium values are found by setting $\frac{dx_i}{dt} = 0$ in (1). They are

$$x_i = \frac{(B+C)I}{A+I} \left(\theta_i - \frac{C}{B+C} \right) . \tag{2}$$

Equation (2) exhibits four main features:

(a) <u>Factorization and Automatic Tuning of Sensitivity</u>

Term $\theta_i - \frac{C}{B+C}$ depends on the ith reflectance θ_i of the input pattern. It is independent of the background intensity I. Formula (2) <u>factorizes</u> information about reflectance from information about background intensity. Due to the factorization property, x_i remains proportional to $\theta_i - \frac{C}{B+C}$ no matter how large I is chosen. In other words, x_i does not saturate.

(b) <u>Adaptation Level, Noise Suppression, and Symmetry-Breaking</u>

Output signals from cell v_i are emitted only if the potential x_i is depolarized. By (1), x_i is depolarized only if term $\theta_i - \frac{C}{B+C}$ is positive. Because the reflectance θ_i must exceed $\frac{C}{B+C}$ to depolarize x_i, term $\frac{C}{B+C}$ is called the <u>adaptation level</u>. The size of the adaptation level depends on the ratio of C to B.

Typically $B \gg C$ in vivo, which implies that $\frac{C}{B+C} \approx 1$. Were not $\frac{C}{B+C} \approx 1$, no choice of θ_i could depolarize the cell since θ_i, being a ratio, never exceeds 1.

The most perfect choice of the ratio of C to B is $\frac{C}{B} = \frac{1}{n-1}$ since then $\frac{C}{B+C} = \frac{1}{n}$. In this case, any uniform input pattern $I_1 = I_2 = \ldots = I_n$ is suppressed by the network because then all $\theta_i = \frac{1}{n}$. Since also $\frac{C}{B+C} = \frac{1}{n}$, all $x_i = 0$ given any input intensity. This property is called noise suppression, or the suppression of zero spatial frequency patterns. Noise suppression guarantees that only nonuniform reflectances of the input pattern can ever generate output signals.

The inequality $B \gg C$ is called a symmetry-breaking inequality for a reason that is best understood by considering the special case when $\frac{C}{B} = \frac{1}{n-1}$. The ratio $\frac{1}{n-1}$ is also, by (1), the ratio of the number of cells excited by input I_i divided by the number of cells inhibited by input I_i. Noise suppression is due to the fact that the asymmetry of the intercellular on-center off-surround interactions is matched by the asymmetry of the intracellular saturation points. In other words, the symmetry of the network as a whole is "broken" to achieve noise suppression. Any imbalance in this matching of intercellular to intracellular parameters will either increase or decrease the adaptation level and thereby modify the noise suppression prope ty.

This symmetry-breaking property of shunting networks leads to a different theory of how on-center off-surround anatomies develop then is implied by an additive theory such as a Fourier or a Laplacian theory, if only because additive theories do not possess excitatory and inhibitory saturation points. In Grossberg (1978a, 1982a) I suggested how the choice of intracellular saturation points in a shunting network may influence the development of intercellular on-center off-surround connections to generate the correct balance of intracellular and intercellular parameters. An incorrect balance could suppress all input patterns by causing

a pathologically large adaptation level. My suggestion is that the balance of intracellular saturation points determines the balance of morphogenetic substances that are produced at the target cells to guide the growing excitatory and inhibitory pathways.

(c) <u>Weber-Law Modulation</u>

Term $\theta_i - \frac{C}{B+C}$ is modulated by the term $(B+C)I(A+I)^{-1}$, which depends only on the background intensity I. This term takes the form of a Weber law (Cornsweet, 1970). Thus (2) describes Weber law modulation of reflectance processing above an adaptation level.

(d) <u>Normalization and Limited Capacity</u>

The total activity of the network is

$$x = \sum_{k=1}^{n} x_k = \frac{[B-(n-1)C]I}{A+I} \quad . \tag{3}$$

By (3), x is independent of the number n of cells in the network if either C=0 or $\frac{C}{B+C} = \frac{1}{n}$. In every case, $x \leq B$ no matter how intense I becomes, and B is independent of n. This tendency for total activity not to grow with n is called total activity <u>normalization</u>. Normalization implies that if the reflectance of one part of the input pattern increases while the total input activity remains fixed, then the cell activities corresponding to other parts of the pattern decrease.

Weber law modulated reflectance processing helps to explain aspects of brightness constancy, whereas the normalization property helps to explain aspects of brightness contrast (Grossberg, 1981). The two types of property are complementary aspects of the same dynamical process.

22. <u>Pattern Matching and Multidimensional Scaling without a Metric</u>

The interaction between reflectance processing and the adaptation level implies that the sum of two mismatched input patterns from two separate input sources will be inhibited by network (1). This is because the mismatched peaks and troughs of

the two input patterns will add to yield an almost uniform total input pattern, which will be quenched by the noise suppression property.

By contrast, the sum of two matched input patterns is a pattern with the same reflectances θ_i as the individual patterns. However the total activity I+J of the summed pattern exceeds the total activities I and J of the individual patterns. Consequently, by (2) the activities in response to the summed pattern are

$$x_i = \frac{(B+C)(I+J)}{A+I+J} \left(\theta_i - \frac{C}{B+C} \right). \tag{4}$$

which exceed the activities in response to the separate patterns. Network activity is hereby amplified in response to matched patterns and attenuated in response to mismatched patterns due to an interaction between reflectance processing, the adaptation level, and Weber law modulation.

The fact that the activity of each cell in a competitive network can depend on how well two input patterns match is of great importance in my theory. Pattern matching is not just a local property of input sizes at each cell. A given cell can receive two different inputs, yet these inputs may be part of perfectly matched patterns, hence the cell activity is amplified. A given cell can receive two identical inputs, yet these inputs may be part of badly mismatched patterns, hence the cell activity is suppressed.

This matching property avoids the homuncular dilemma by being an automatic consequence of the network's pattern registration process. Various models in Artificial Intelligence, by contrast, use a Euclidian distance $\sum_{k=1}^{n} (I_k - J_k)^2$ or some other metric to compute pattern matches (Klatt, 1980; Newell, 1980). Such an approach requires a separate processor to compute a scalar distance between two patterns before deciding how to tack the results of this scalar computation back onto the mainstream of computational activity. A metric also misses properties of the competitive matching process which are crucial in the study of spatial vision, as well as in other pattern recognition problems wherein multiple scales are needed to unambiguously represent the data.

In the competitive matching process, a match not only encodes the matched pattern; it also amplifies this pattern. A metric does not encode a pattern because it is a scalar, rather than a vector. A metric does not amplify the matched patterns because it is minimized rather than maximized by a pattern match. Moreover, what is meant by matching differs in a metric than in a shunting network. A metric makes local matches between corresponding input intensities, whereas a network matches reflectances, which depend upon the entire pattern. One could of course use a metric to match ratios of input intensities, but this computation requires an extra homuncular processing step and is insensitive to overall input intensity, which is not true of the network matching mechanism.

Although the properties of metric matches are disappointing in comparison to properties of feedforward network matching, they are totally inadequate when compared to properties of feedback network matching. In a feedback context, network matching has hysteresis properties which can maintain a match during slow deformations of the input patterns, and pattern completion properties which can deform an approximate match into a better "fused" match (Grossberg, 1980a).

The primary use of network matching in my binocular theory is to show how those spatial scales which achieve the best binocular match of monocular data from the two eyes can resonate energetically, whereas those spatial scales which generate a mismatched binocular interpretation of the monocular data are energetically attenuated. The ease with which these multidimensional scaling effects occur is due to properties that obtain in even the simplest competitive networks. I use the term "multidimensional scaling" deliberately, since similar competitive rules often operate on a higher perceptual and cognitive level (Grossberg, 1978a), where metrical concepts have also been used as explanatory tools (Osgood, Suci, and Tannenbaum, 1957; Shepard, 1980).

An inadequate model of how cell activity reflects matching can limit a theory's predictive range. For example, in a binocular context, I will use this relationship to explain several types of data, including the coexistence of Fechner's paradox

and binocular brightness summation (Blake, Sloane, and Fox, 1981), and the choice

between binocular fusion and rivalry within a given spatial scale (Kaufman, 1974;

Kulikowski, 1978). A reason for binocular brightness summation is already evident

in equation (4). The effects of activities I and J on x_i exceed those expected

from noninteracting independent detectors, but are less than the sum I+J, as a

result of Weber law modulation (Cogan, Silverman, and Sekuler, 1982).

23. Weber Law and Shift Property without Logarithms

The simple equation (1) has other properties which are worthy of note. These

properties describe other aspects of how the network retunes itself in response

to changes in background activity.

The simplest consequence of this retuning property is the classical Weber

law

$$\frac{\Delta I}{I} = \text{constant}, \tag{5}$$

where ΔI is the just noticeable increment above a background intensity I. The

approximate validity of (5) has encouraged the belief that logarithmic processing

determines visual sensitivity (Cornsweet, 1970; Land, 1977), since $\Delta \log I = \frac{\Delta I}{I}$,

despite the fact that the logarithm exhibits unphysical infinities at small and

large values of its argument. In fact, Cornsweet (1970) built separate theories of

reflectance processing and of brightness perception by using logarithms to discuss

reflectances and shunting functions like $I(A+J)^{-1}$ to discuss brightness. By contrast,

shunting equations like (2) join together reflectance processing and brightness pro-

cessing into a single computational framework.

Power laws have often been used in psychophysics instead of logarithms (Stevens,

1959). It is therefore of interest that equation (2) guarantees reflectance proces-

sing undistorted by saturation if the inputs I_i are power law outputs $I_i = \bigvee J_i^p$ of

the activities J_i at a prior processing stage. Reflectance processing is preserved

under power law transformations because the form of (2) is left invariant by such a

transformation. In particular,

$$x_i = \frac{B\theta_i I}{A^* + I} \qquad (6)$$

where

$$\theta_i = \frac{\phi_i^p}{\sum\limits_{k=1}^{n} \phi_k^p} \quad , \qquad (7)$$

$$I = J^p \quad , \qquad (8)$$

and

$$A^* = A\lambda^{-p} \left(\sum\limits_{k=1}^{n} \phi_k^p\right)^{-1} . \qquad (9)$$

To show how the Weber law (5) approximately obtains in (2), choose

$$I_1 = K + \Delta I, \text{ and } I_2 = I_3 = \ldots = I_n = K. \qquad (10)$$

Then the total input before *increment* ΔI is applied to I_1 is $I = nK$. By (2)

$$x_1 = \frac{(B+C)(I+\Delta I)}{A+I+\Delta I} \left(\frac{K+\Delta I}{nK+\Delta I} - \frac{C}{B+C} \right) . \qquad (11)$$

If $I \gg \Delta I$ and $n \gg 1$, then

$$\frac{K+\Delta I}{nK+\Delta I} - \frac{C}{B+C} = \frac{\Delta I}{I} \frac{(n-1)}{n} \frac{I}{I+\Delta I} + D$$

$$\cong \frac{\Delta I}{I} + D \qquad (12)$$

where $D = \frac{1}{n} - \frac{C}{B+C}$. If $I \gg A$, then

$$\frac{(B+C)(I+\Delta I)}{A+I+\Delta I} \cong B+C \qquad (13)$$

Consequently

$$x_1 \cong (B+C)(\frac{\Delta I}{I} + D). \tag{14}$$

If x_1 is detectable when it exceeds a threshold Γ, then

$$\frac{\Delta I}{I} \cong W \tag{15}$$

where

$$W = \frac{\Gamma}{B+C} - D = \text{constant} . \tag{16}$$

A more precise version of the Weber law (5) is the shift property. This property says that the region of maximal visual sensitivity shifts without compression as the background off-surround intensity is parametrically increased (Werblin, 1971). The shift property obtains when the on-center input I_i is plotted in logarithmic coordinates despite the fact that (2) does not describe logarithmic processing.

The shift property is important in a multidimensional parallel processing framework wherein changes in the number and intensity of active input sources can fluctuate wildly through time. Given the shift property, one can fix the activity scale (-C, B) and the network's output threshold once and for all without distorting the network's decision rules as the inputs fluctuate through time. A fixed choice of operating range and of output thresholds is impossible in a multidimensional parallel processing theory that is built up from additive processors. If a fixed threshold is selective when m converging input channels are active, then it may not generate any outputs whatsoever when n << m input channels of comparable intensity are active, and may unselectively generate outputs whenever n >> m input channels are active. Such a theory needs to continually redefine how big its thresholds should be as the input load fluctuates through time.

To derive the shift property, rewrite (2) as

$$x_i = \frac{(B+C)I_i - CI}{A+I} . \tag{17}$$

Also write I_i in logarithmic coordinates as $M = \log_e I_i$, or $I_i = e^M$, and the total

off-surround input as $L = \sum_{k=1} L_k$. Then in logarithmic coordinates, (17) becomes

$$x_i(M,L) = \frac{Be^M - CL}{A+L+e^M} . \qquad (18)$$

The question of shift invariance is: Does there exist a shift S such that

$$x_i(M+S,L_1) \equiv x_i(M,L_2). \qquad (19)$$

for all M, where S depends only on L_1 and L_2? The answer is yes if C=0 (no hyper-polarization). Then

$$S = \log_e \left(\frac{A+L_1}{A+L_2} \right) \qquad (20)$$

which shows that successively increasing L by linear increments ΔL in (18) causes

progressively smaller shifts S in (20). In particular, if $L_1 = (n+1)\Delta L$ and

$L_2 = n\Delta L$, then S approaches zero as n approaches infinity. If $C > 0$, then (19)

implies that

$$S = \log_e \left[\frac{AB + (B+C)L_1 + AC(L_1-L_2)e^{-M}}{AB + (B+C)L_2} \right] \qquad (21)$$

By (21), S depends on M only via term $AC(L_1-L_2)e^{-M}$, which rapidly decreases

as M increases. Thus the shift property improves, rather than deteriorates, at

the larger intensities M which might have been expected to cause saturation.

Moreover, if $B \gg C$, as occurs physically, then (20) is approximately valid at

all values of $M \geq 0$.

24. Edge, Spatial Frequency, and Reflectance Processing by the Receptive Fields

    of Distance-Dependent Feedforward Networks

    Equation (1) is based on several assumptions which do not always occur in vivo.

It is the task of the mathematical classification theory to test the consequences of

modifying these assumptions. One such assumption says that the inhibitory inputs

excite all off-surround channels with equal strength, as in term $-(x_i+C) \sum\limits_{k \neq i} I_k$ of (1). Another assumption says that only the ith channel is excited by the ith input, as in term $(B-x_i)I_i$ of (1). In a general feedforward shunting network, both the excitatory and the inhibitory inputs can depend on the distance between cells, as in the feedforward network

$$\frac{dx_i}{dt} = -Ax_i+(B-x_i) \sum_{k=1}^{n} I_k D_{ki} - (x_i+C) \sum_{k=1}^{n} I_k E_{ki} \ . \tag{22}$$

Here the coefficients D_{ki} and E_{ki} describe the fall-off with the distance between cells v_k and v_i of the excitatory and inhibitory influences, respectively, of input I_k on cell v_i.

Equation (22) exhibits variants of all the properties enjoyed by equation (1). These properties follow from the equilibrium activities of (22), namely

$$x_i = \frac{F_i I}{A + G_i I} \tag{23}$$

where

$$F_i = \sum_{k=1}^{n} \theta_k \ (BD_{ki} - CE_{ki}) \tag{24}$$

and

$$G_i = \sum_{k=1}^{n} \theta_k \ (D_{ki} + E_{ki}) \ , \tag{25}$$

in response to a sustained input pattern $I_i = \theta_i I$, $i = 1,2,\ldots\ldots n$. See Ellias and Grossberg (1975) and Grossberg (1981) for a discussion of these properties. For present purposes, I will focus on the fact that the noise suppression property in the network (22) implies an edge detection and spatial frequency detection capability in addition to its pattern matching capability.

The noise suppression property in (23) is guaranteed by imposing the inequalities

$$B \sum_{k=1}^{n} D_{ki} \leq C \sum_{k=1}^{n} E_{ki} \ , \tag{26}$$

$i = 1, 2, \ldots .n$. Noise suppression follows from (26) because then all $x_i \leq 0$ in response to a uniform pattern (all $\theta_i = \frac{1}{n}$) by (23) and (24). The inequalities (26) say, just as in Section 21, that there exists a matched symmetry-breaking between the spatial bandwidths of excitatory and inhibitory intercellular signalling and the choice of inhibitory and excitatory intracellular saturation points $-C$ and B, respectively.

A distance-dependent network with the noise suppression property can detect edges and other nonuniform spatial gradients for the following reason. Those cells which perceive a uniform input pattern within the breadth of their excitatory and inhibitory scales are suppressed by the noise suppression property no matter how intense the pattern activity is (Figure 7). Only those cells which perceive a non-

Figure 7

uniform pattern with respect to their scales can generate suprathreshold activity. This is also true in a suitably designed additive network (Ratliff, 1965).

When the interaction coefficients D_{ki} and E_{ki} of (22) are Gaussian functions of distance, as in $D_{ki} = D \exp [-\mu(k-i)^2]$ and $E_{ki} = E \exp [-\nu(k-i)^2]$, then the equilibrium activities x_i in (23) modify and generalize the model of receptive field properties that is currently used to fit a variety of visual data. In particular, the term F_i in (24) that appears in the numerator of x_i depends on sums of differences of Gaussians. Difference-of-Gaussian form factors for studying receptive field responses appear in the work of various authors (Blakemore, Carpenter, and Georgeson, 1970; Ellias and Grossberg, 1975; Enroth-Cugell and Robson, 1966; Levine and Grossberg, 1976; Rodieck and Stone, 1965; Wilson and Bergen, 1979). At least two properties of (23) can distinguish it from an additive difference-of-Gaussian theory. The first property is that F_i in (24) depends on weighted difference of Gaussian factors $BD_{ki} - CE_{ki}$, such that the weights B and $-C$ equal the excitatory and inhibitory saturation points, respectively. Consequently, given fixed sizes of $D_{ii} = D$ and $E_{ii} = E$ and the noise suppression property, if the symmetry-breaking inequality $B >> C$

holds then the ratio $\mu\nu^{-1}$ of excitatory to inhibitory spatial bandwidths should be larger in a shunting theory than in an additive theory.

A second way to experimentally distinguish between additive and shunting receptive field models is to test whether the contrast of the patterned responses changes as a function of suprathreshold background luminance. In an additive theory, the answer is "no". In a distance-dependent shunting equation such as (23), the answer is "yes"(Ellias and Grossberg, 1975; Grossberg, 1981). The ratios which determine x_i in (23) lead to changes of contrast as the background intensity I increases only because the coefficients D_{ki} and E_{ki} are distance-dependent. In a shunting network with a very narrow excitatory bandwidth and a very broad inhibitory bandwidth, the relative sizes of the x_i are independent of I. The contrast changes which occur as I increases in the distance-dependent case can be viewed as a partial breakdown of reflectance processing at high I levels due to the inability of inhibitory gain control to fully compensate for saturation effects.

The edge enhancement property of a *feedforward competitive* network confronts us with the full force of the filling-in dilemma. If only edges can be detected by a network once it is constrained to satisfy, even approximately, such a basic property as noise suppression, then how do we fill-in the interiors of extended bodies?

25. Statistical Analysis by Structural Scales: Edges with Scaling and Reflectance Properties Preserved

Before facing this dilemma, I need to review other properties of the excitatory input term $\sum_{k=1}^{n} I_k D_{ki}$ and the inhibitory input term $\sum_{k=1}^{n} I_k E_{ki}$ in (22). Let the interaction coefficients D_{ki} and E_{ki} be distance-dependent, so that $D_{ki} = D(\,k-i\,)$ and $E_{ki} = E(|k-i|)$ where the functions $D(j)$ and $E(j)$ are decreasing functions of j, such as Gaussians. Then the input terms $\sum_{k=1}^{n} I_k D_{ki}$ cross-correlate the input pattern $(I_1, I_2, \ldots, I_n)$ with the kernel $D(j)$. Similarly, the input terms $\sum_{k=1}^{n} I_k E_{ki}$ cross-

correlate the input pattern $(I_1, I_2, \ldots\ldots, I_n)$ with the kernel $E(j)$. These statistics of the input pattern, rather than the input pattern itself, are the local data to which the network reacts. I will call the kernels $D(j)$ and $E(j)$ structural scales of the network to distinguish them from the functional scales that will be defined below. The structural scales perform a statistical analysis of the data before the shunting dynamics further transform these data statistics. Although terms like $\sum_{k=1}^{n} I_k D_{ki}$ are linear functions of the inputs I_k, the inputs are themselves often nonlinear functions, notably S-shaped or sigmoidal functions, of outputs from prior network stages (Section 28). Thus the statistical analysis of input patterns is in general nonlinear.

These concepts are elementary, as well as insufficient for our purposes. It is, however, instructive to review how statistical preprocessing of an input pattern influences the network's reaction to patterns more complex than a rectangle, say a periodic pattern of high spatial frequency bars superimposed on a periodic

Figure 8

pattern of low spatial frequency bars (Figure 8a). Suppose for definiteness that the excitatory scale $D(j)$ is narrower than the inhibitory scale $E(j)$ to prevent the occurrence of spurious peak splits and multiple edge effects in the network's response to spots and bars of input (Ellias and Grossberg, 1975). Then the excitatory structural bandwidth determines a unit length over which input data is statistically pooled, whereas the inhibitory structural bandwidth determines a unit length over which the pooled data of nearby populations are evaluated for their uniformity.

A network whose excitatory bandwidth approximates width λ can react to the input pattern with a periodic series of smoothed bumps (Figure 8b). By contrast, a

network whose excitatory bandwidth equals period 2α but is less than the entire
pattern width reacts only to the smoothed edges of the input pattern (Figure 8c).
The interior of the input pattern is <u>statistically uniform</u> with respect to the
larger structural scale, and therefore its interior is inhibited by noise suppres-
sion. As the excitatory bandwidth increases further, the smoothed edges are lumped
together until the pattern generates a single centered hump, or spot, of network
activity (Figure 8d). This example illustrates how the interaction of a broad struc-
tural scale with the noise suppression mechanism can inhibit all but the smoothed
edges of a finely and regularly textured input pattern. After inhibition takes place,
the spatial breadth of the surviving edges responses depends on both the input tex-
ture and the structural scale; the edges have not lost their scaling properties. The
peak height of these edges responses compute a measure of the pattern's reflectances
near its boundary, since ratios of input intensities across the network determine
the steady-state potentials x_i in (23). Rather than discard these monocular scaling
and lightness properties, as in a zero-crossing computation, I will use them in an
essential way below as the data with which to build up binocular resonances.

26. Correlation of Monocular Scaling with Binocular Fusion

The sequence of activity patterns in Figure 8b,c,d is reversed when an observer
steadily approaches the picture in Figure 8a. Then the spot in Figure 8d bifurcates
into two boundary responses, which in turn bifurcate into a regular pattern of
smoothed bumps, which finally bifurcate once again to reveal the high frequency com-
ponents within each bump. If the picture starts out sufficiently far away from the
observer, then the first response in each of the observer's spatial scales is a
spot, and the bifurcations in the spot will occur in the same order. However, the
distance at which a given bifurcation occurs depends on the spatial scale in ques-
tion. Other things being equal, a prescribed bifurcation will occur at a greater
distance if the excitatory bandwidth of the spatial scale is narrower (high spatial
frequency). Furthermore, the registration of multiple spatial frequencies, or even

of multiple spots, in the picture will not occur in a spatial scale whose excitatory bandwidth is too broad (low spatial frequency).

The same sequence of bifurcations can occur within the multiple spatial scales corresponding to each eye. If the picture is simultaneously viewed by both eyes, the question naturally arises: How do the two activity patterns within each monocular scale binocularly interact at each distance? Let us assume for the moment, as in the Kaufman (1974) and Kulikowski (1978) experiments, that as the disparity of two monocular patterns increases, it becomes harder for the high spatial frequency scales to fuse them. Since disparity decreases with increasing distance, all scales can binocularly fuse their respective patterns (supposing they are detectable at all) when the distance is great enough, but the lower spatial frequency scales can maintain fusion over a broader range of decreasing distances than can the higher spatial frequency scales. Other things being equal, the scales which can most easily binocularly fuse their two monocular representations of a picture at a given distance are the scales which average away the finer features in the picture. It therefore seems natural to ask: Does the broad spatial smoothing within low spatial frequency scales enhance their ability to binocularly fuse disparate monocular activity patterns?

Having arrived at this issue, we now need to study those properties of feedback competitive shunting networks that will be needed to design scale-sensitive binocular resonances in which the fusion event is only one of a constellation of interrelated depth, length, and lightness properties.

27. Noise Suppression in Feedback Competitive Networks

The noise-saturation dilemma confronts all cellular tissues which process input patterns, whether the cells exist in a feedforward or in a feedback anatomy. As part of the mathematical classification theory, I will therefore consider shunting interactions in a feedback network wherein excitatory signals are balanced by inhibitory signals. Together these feedback signals are capable of returning network sensitivity in response to fluctuating background activity levels.

The feedback analog of the distance-dependent feedforward network (22) is

$$\frac{dx_i}{dt} = -Ax_i + (B-x_i) \left[J_i + \sum_{k=1}^{n} f(x_k)D_{ki} \right]$$

$$- (x_i + C) \left[K_i + \sum_{k=1}^{n} g(x_k)E_{ki} \right], \tag{27}$$

$i = 1,2,\ldots\ldots n$. As in (22), term $-Ax_i$ describes the spontaneous decay of activity at rate $-A$. Term $(B-x_i)J_i$ describes the excitatory effect of the feedforward excitatory input J_i, which was chosen equal to $\sum_{k=1}^{n} I_k D_{ki}$ in (22). Term $-(x_i + C)K_i$ is also a feedforward term due to inhibition of activity by the feedforward inhibitory input K_i, which was chosen equal to $\sum_{k=1}^{n} I_k E_{ki}$ in (22). The new excitatory feedback term $\sum_{k=1}^{n} f(x_k)D_{ki}$ describes the total effect of all the excitatory feedback signals $f(x_k)D_{ki}$ from the cells v_k to v_i. The function $f(x_i)$ transmutes the activity, or potential, of x_i into a feedback signal $f(x_i)$, which can be interpreted either as a density of spikes per unit time interval, or as an electrotonic influence, depending on the situation. The inhibitory feedback term $\sum_{k=1}^{n} g(x_k)E_{ki}$ determines the total effect of all the inhibitory feedback signals $g(x_k)E_{ki}$ from the cells v_k to v_i. As in (22), the interaction coefficients D_{ki} and E_{ki} are often defined by kernels $D(j)$ and $E(j)$, such that $E(j)$ decreases more slowly than $D(j)$ as a function of increasing values of j.

The problem of noise suppression is just as basic in feedback networks as in feedforward networks. Suppose, for example, that the feedforward inputs and the feedback signals both use the same interneurons and the same statistics of feedback signalling ($f(x_i) = g(x_i)$) to distribute their values across the network. Then (27) becomes

$$\frac{dx_i}{dt} = -Ax_i + (B - x_i) \sum_{k=1}^{n} [I_k + f(x_k)] \ D_{ki}$$

$$- (x_i + C) \sum_{k=1}^{n} [I_k + f(x_k)] E_{ki} \tag{28}$$

$i = 1,2,\ldots\ldots,n$. In such a network, ᴛ · same criterion of uniformity is applied both to feedforward and to feedback signals. Both processes share the same structural scales. Correspondingly, in (28) as in (22) the single inequality

$$B \sum_{k=1}^{n} D_{ki} \leq C \sum_{k=1}^{n} E_{ki} \qquad (26)$$

suffices to suppress both uniform feedforward patterns and uniform feedback patterns.

28. Sigmoid Feedback Signals and Tuning

Another type of noise suppression is also needed for a feedback network to function properly. This is true because certain positive feedback functions $f(w)$ can amplify even very small activities w into large activities. Noise amplification can flood the network with internally generated noise capable of massively distorting the processing of feedforward inputs. Pathologies of feedback signalling have been suggested to cause certain seizures and hallucinations (Ellias and Grossberg, 1975; Grossberg, 1973; Kaczmarek and Babloyantz, 1977).

In Grossberg (1973), I proved as part of the mathematical classification theory that the simplest physically plausible feedback signal which is capable of attenuating, rather than amplifying, small activities is a sigmoid, or S-shaped, signal function (Figure 9). Several remarks should be made about this result.

Figure 9

The comment is sometimes made that you only need a signal threshold to prevent noise amplification (Figure 9). This is true but insufficient, because a threshold signal function does not perform the same pattern transformation as a sigmoid signal function. For example, in a shunting network with a narrow on-center and a broad off-surround, a threshold signal chooses the population that receive the largest input for activity storage and suppresses the activities of all other populations. By contrast, a sigmoid signal implies the existence of a quenching threshold (QT). This means that the activities of populations whose initial activation is less than the QT are suppressed, whereas the activity pattern of populations whose initial activities exceed the QT is contrast enhanced before being stored. I identify this storage process with storage in short term memory (STM). In a network that possesses a QT, any operation which alters the QT can sensitize or desensitize the network's ability to store input data (Figure 10). This tuning

Figure 10

property is trivialized in a network that chooses the population which receives the largest input for STM storage.

Another important point is that the QT does not equal the turning point, or manifest threshold, of the sigmoid signal function (Figure 9). The QT depends on the turning point, on the slope of the signal function, on the number of excitable sites, on the geometry of intercellular feedback signalling via the coefficients D_{ki} and E_{ki}, etc. This fact must be understood to effectively argue that the breakdown of any of several mechanisms can induce seizures or hallucinations by causing the QT to assume abnormally small values.

29. The Interdependence of Contrast Enhancement and Tuning

The existence of a QT suggests that the contrast enhancement of input patterns that is ubiquitous in the nervous system is not an end in itself (Ratliff, 1965). In feedback competitive shunting networks, contrast enhancement is a mathematical consequence of the noise suppression property. This fact is emphasized by the observation that linear feedback signals can perfectly store an input pattern's reflectances - in particular, do not enhance the pattern - but only at the price of amplifying network noise (Grossberg, 1973, 1981). Contrast enhancement by a feedback network in its suprathreshold activity range follows from noise suppression by the network in its subthreshold activity range. Contrast enhancement can intuitively be understood if a feedback competitive network possesses a normalization property like that of a feedforward competitive network (Section 21). If small activities are attenuated by noise suppression and total activity is approximately conserved due to normalization, then large activities will be enhanced.

30. Normalization in a Feedback Competitive Network: A Limited Capacity Short
    Term Memory System

Suitably designed feedback competitive networks do possess a normalization property. Recall from Section 21 that in a feedforward competitive network, the total activity can increase with the total input intensity but is independent of the number of active cells. This is true only if the inhibitory feedforward interaction ι of long range across the network cells. If the strengths of the inhibitory pathways are weakened or fall off rapidly with distance, then the normalization property is weakened also, and saturation can set in at high input intensities. The same property tends to hold for the feedforward terms $(B-x_i)J_i$ and $-(x_i + C)K_i$ of (27).

The normalization property of a feedback competitive network is more subtle (Grossberg, 1973, 1981). If such a network is excited to suprathreshold activities

and if the exciting inputs are then terminated, then the total activity of
the network approaches one of perhaps several positive equilibrium values,
all of which tend to be independent of the number of active cells. Thus if
the activity of one cell is for some reason increased, then the activities
of other cells will decrease to satisfy the normalization constraint unless
the system as a whole is attracted to a different equilibrium value. This
limited capacity constraint on short term memory is an automatic property
in our setting. It is postulated without mechanistic justification in various
other accounts of short term memory processing (Raaijmakers and Shiffrin,
1981, p. 126).

31. Propagation of Normalized Disinhibitory Crests

Just as in feedforward networks, the feedback normalization property is
weakened if the inhibitory path strengths are chosen to decrease more rapidly
with distance. Then the normalization property tends to hold among subsets of
cells that lie within one bandwidth of the network's inhibitory structural
scale. In particular, if some cell activities are enhanced by a given amount,
then their neighbors will tend to be suppressed by a comparable amount. The
neighbors of these neighbors will then be enhanced by a similar amount, and so
on. In this way, a disinhibitory wave can propagate across a network in such a
way that each crest of the wave inherits, or "remembers", the activity of the
previous crest. This implication of the normalization property in a feedback
network with finite structural scales will be important in my account of
filling-in. Normalization within a structural scale also imparts the network's
activity patterns with constancy and contrast patterns, as in the case of
feedforward competitive networks (Section 24). In a feedback context, however,
constancy and contrast properties can propagate far beyond the confines of a
single structural scale because of normalized disinhibitory properties such as
Figure 11 depicts.

32. Structural vs. Functional Scales

The propagation process depicted in Figure 11 needs to be understood in greater detail because it will be fundamental in all that follows. A good way to approach this understanding is to compare the reactions of competitive feedforward networks with those of competitive feedback networks to the same input patterns.

Let us start with the simplest case. Choose C = 0 in (22) and (27). This prevents the noise suppression inequalities (26) from holding. Although feedforward and feedback inhibition are still operative, activities cannot be inhibited below zero in this case. Consequently, a uniform input pattern can be attenuated but not entirely suppressed. Choose a sigmoidal feedback signal function to prevent noise amplification, and thus to contrast enhance the pattern of suprathreshold activities. These hypotheses enable us to study the main effects of feedback signalling unconfounded by the effect of noise suppression.

What happens when we present a rectangular input pattern (Figure 11a) to

Figure 11

both networks? Due to the feedforward inhibition in (22), the *feedforward* network enhances the edges of the rectangle and attenuates its interior (Figure 11b). By contrast, the feedback network elicits a regularly spaced series of excitatory peaks across the cells that receive the rectangular input (Figure 11c). This type of reaction occurs even if the input pattern is not contrast-enhanced by a feedforward inhibitory stage, as in Figure 11b, before feedback inhibition can act on the contrast-enhanced pattern. The pattern of Figure 11c is elicited even if the feedback acts directly on the rectangular input pattern (Ellias and Grossberg, 1975).

The spatial bandwidth between successive peaks in Figure 11c is called the *functional scale* of the feedback network. My first robust points are that a functional scale can exist in a feedback network but not in a feedforward network, and that, although the functional scale is related to the structural scale of a

feedback network, the two scales are not identical. I will discuss the functional

scale given $C = 0$ before reinstating the noise suppression inequalities (26)

because the interaction between contrast enhancement and noise suppression in a

feedback network is a much more subtle issue.

33. Disinhibitory Propagation of Funtional Scaling from Boundaries to Interiors

To see how a functional scale develops, let us consider the network's response

to the rectangular input pattern on a moment-to-moment basis. All the populations

v_m that are excited by the rectangle initially receive equal inputs. All the acti-

vities x_m of these populations therefore start to grow at the same rate. This growth

process continues until the feedback signals $f(x_m)D_{mi}$ and $g(x_m)E_{mi}$ can be registered

by the other populations v_i. Populations v_i which are near the rectangle's boundary

receive smaller total inhibitory signals $\sum_{m=1}^{n} g(x_m)E_{mi}$ than populations which lie near-

er to the rectangle's center, even when all the rectangle-excited activities x_m are

equal. This is because the interaction strengths $E_{mi} = E(|m-i|)$ are distance-depen-

dent, and the boundary populations receive no inhibition from contiguous populations

that lie outside the rectangle.

As a result of this inhibitory asymmetry, the activities x_i near the boundary

start to grow faster than contiguous activities x_j nearer to the center. The inhi-

bitory feedback signal $g(x_i)E_{ij}$ from v_i to v_j begins to exceed the inhibitory feed-

back signal $g(x_j)E_{ji}$ from v_j to v_i, because $x_i > x_j$ and $E_{ij} = E_{ji}$. Thus although all

individual feedback signals among rectangle-excited populations start

out equal, they are soon differentiated due to a second-order effect whereby the

boundary bias in the spatial distribution of the total inhibitory feedback signals

is mediated by the activities of individual populations.

As the interior activities x_j get differentially inhibited, their inhibitory

signals $g(x_j)E_{jk}$ to populations v_k which lie even deeper within the rectangle's

interior become smaller. Now the total pattern of inputs plus feedback signals is

no longer uniform across the populations v_j and v_k. The populations v_k are favored.

Contrast enhancement bootstraps their activities x_k to larger values. Now these

populations can more strongly inhibit neighboring populations that lie even deeper

into the rectangle's interior, and the process continues in this fashion.

The boundary asymmetry in the total inhibitory feedback signals hereby propa-

gates ever deeper into the rectangle's interior by a process of distance-dependent

disinhibition and contrast enhancement until all the rectangle-excited populations

are filled-in by a series of regularly spaced activity peaks as in Figure 11c.

34. Quantization of Functional Scales: Hysteresis and Uncertainty

As I mentioned in Section 32, two distinct types of spatial scales can be dis-

tinguished in a feedback network. The structural scales $D(j)$ and $E(j)$ describe how

rapidly the network's feedback interaction coefficients decrease as a function of

distance. The functional scale describes the spatial wavelength of the disinhibi-

tory peaks that arise in response to prescribed input patterns. Although these two

types of scale are related, they differ in fundamental ways.

They are related because an increase in a network's structural scales can

cause an increase in the functional scale with which it fills-in a given input

pattern (Ellias and Grossberg, 1975). This is due to two effects acting together.

A slower decrease of $D(j)$ with increasing distance j can increase the number of con-

tiguous populations that pool excitatory feedback. This effect can broaden the peaks

in the activity pattern. A slower decrease of $E(j)$ with increasing distance j can

increase the number of contiguous populations which can be inhibited by an activity

peak. This effect can broaden the troughs in the activity pattern. This relationship between structural scales and functional scales partially supports the intuition that visual processing includes a spatial frequency analysis of visual data (Graham, 1981; Robson, 1976), because if several feedback networks with distinct structural scales received the same input pattern, then they will each generate distinct functional scales such that smaller structural scales tend to generate smaller functional scales. However, the functional scale does not equal the structural scale, and its properties represent a radical departure from feedforward linear ideas.

The most important of these differences can be summarized as follows. The functional scale is a _quantized_ property of the _interaction_ between the network and _global_ features of an input pattern, such as its length. Unlike a structural scale, a functional scale is not just a property of the network. Nor is it just a property of the input pattern. The interaction between pattern and network literally creates the functional scale. The quantized nature of this interaction is easy to state because it is so fundamental. (The reader who knows some quantum theory, notably Bohr's original model of the hydrogen atom, might find it instructive to compare the two types of quantization.)

The length L of a rectangular input pattern might equal a nonintegral multiple of a network's structural scales, but there obviously can exist only a integral number of disinhibitory peaks in the activity pattern induced by the rectangle. The feedback network therefore quantizes its activity in a way that depends on the global structure of the input pattern. The functional scales must change to satisfy the quantum property as distinct patterns perturb the network, even though the network's structural scales remain fixed.

For example, rectangular inputs of length L, L + ΔL, L + 2ΔL,..., L + ΔL

might all induce M_L peaks in the network's activity pattern. Not until a rectangle

of length L + (ω+1)ΔL is presented might the network respond with M_L + 1 peaks.

This length quantization property suggests a new reason why a network, and percep-

tion (Fender and Julesz, 1967), can exhibit hysteresis as an input pattern is slow-

ly deformed through time. Another consequence of the quantization property is that

the network cannot distinguish certain differences between input patterns. Quanti-

zation implies a certain degree of perceptual uncertainty.

35. Phantoms

The reader might by now have entertained the following objection to these ideas.

If percepts really involve spatially regular patterned responses even to uniform in-

put regions, then why don't we easily see these patterns? I suggest that we some-

times do, as when spatially periodic visual phantoms can be seen superimposed upon

otherwise uniform, and suprisingly large, regions (Smith and Over, 1979; Tynan and

Sekuler, 1975; Weisstein, Maguire, and Berbaum, 1977). The disinhibitory filling-in

process clarifies how these phantoms can cover regions which excite a retinal area

much larger than a single structural scale. I suggest that we do not more often see

phantoms for three related reasons.

During day-to-day visual experience, several functional scales are often simul-

taneously active. The peaks of higher spatial frequency functional scales can over-

lay the spaces between lower spatial frequency functional scales. Retinal tremor

and other eye movements can randomize the spatial phases of, and thereby spatially smooth, the higher frequency scales across the lower frequency scales through time. Even within a single structural scale, if the boundary of an input pattern curves in two dimensions, then the disinhibitory wavelets can cause interference patterns as they propagate into the interior of the activity pattern along rays perpendicular to each boundary element. These interference patterns can also obscure the visibility of a functional scale. Such considerations clarify why experiments in which visual phantoms are easily seen usually use patterns that selectively resonate with a low spatial frequency structural scale that varies in only one spatial dimension.

An important issue concerning the perception of phantoms is whether they are, of necessity, perceivable only if moving displays are used, or whether the primary effect of moving a properly chosen spatial frequency at a properly chosen velocity is to selectively suppress all but the perceived spatial wavelength via noise suppression. The latter interpretation is compatible with an explanation of spatial frequency adaptation using properties of shunting feedback networks (Grossberg, 1980a, Section 12).

A possible experimental approach to seeing functional scales using a stationary display takes the form of a two-stage experiment. First adapt out the high spatial frequencies using a spatial frequency adaptation paradigm. Then fixate a bounded display which is large enough and is shaped properly to strongly activate a low spatial frequency scale in one dimension, and which possesses a uniform interior that can energize periodic network activity.

36. Functional Length and Emmert's Law

Two more important properties of functional scales are related to length and lightness estimates. The functional wavelength defines a length scale. To understand what I mean by this, let a rectangular input pattern of fixed length L excite networks with different structural scales. I hypothesize that the apparent length of the rectangle in each network will depend on the functional scale generated therein. Since a broader structural scale induces a broader functional scale, the activity pattern in such a

network will contain fewer active functional wavelengths. I suggest that this property is associated with an impression of a shorter object, despite the fact that L is fixed.

The reader might object that this property implies too much. Why can a monocularly viewed object have ambiguous length if it can excite a functional scale? I will suggest that under certain, but not all, monocular viewing conditions, an object may excite all the structural scales of the observer. When this happens, the object's length may seem ambiguous. I will also suggest in Section 39 how binocular viewing of a nearby object can selectively excite structural scales which subserve large functional scales, thereby making the object look shorter. By contrast, binocular viewing of a far-away object can selectively excite structural scales which subserve small functional scales, thereby making the object look longer. Thus the combination of binocular selection of structural scales that vary inversely with an object's distance, along with the inverse variation of length estimates with functional scales, can provide an explanation of Emmert's law.

This view of the correlation between perceived length and perceived distance does not imply that the relationship should be veridical - and indeed it sometimes is not (Hagen and Teghtsocnian, 1981) - for the following reasons. The functional scale is a quantized collective property of a nonlinear feedback network rather than a linear ruler. The selection of which structural scales will resonate to a given object and of which functional scales will be generated within these structural scales depend on the interaction with the object in different ways; for one, the choice of structural scale does not depend on a filling-in reaction.

These remarks indicate a sense in which functional scales define an "intrinsic metric," which is independent of cognitive influences but on whose shoulders correlations with motor maps, adaptive chunking and learned feedback expectancy computations can build (Grossberg, 1978a, 1980a). This intrinsic metric helps to explain how monocular scaling effects, such as those described in Section 5, can occur. Once the relevance of the functional scale concept to metrical estimates is broached, one can begin to appreciate how a dynamic "tension" or "force field" or "curved metric" can be generated

whereby objects which excite one part of the visual field can influence the perception of objects at distinct visual positions (Koffka, 1935; Watson, 1978). As we proceed, I will argue that the functional scale concept explicates a notion of dynamic field interactions that escapes the difficulties faced by the Gestaltists in their pioneering efforts to explain global visual interactions.

37. Functional Lightness and the Cornsweet Effect

The functional scale concept clarifies how object boundaries can determine the lightness of object interiors, as in the Cornsweet effect. Other things being equal, a more intense pattern edge will cause larger inhibitory troughs around itself. The inhibitory trough which is interior to the pattern will hereby create a larger disinhibitory peak due to pattern normalization within the structural scale. This disinhibitory process continues to penetrate the pattern in such a fashion that all the interior peak heights are influenced by the boundary peak height because each inhibitory trough "remembers" the previous peak height. The sensitivity of filled-in interior peak size to boundary peak size helps to explain the Cornsweet effect (Section 11).

Crucial to this type of explanation is the idea that the disinhibitory filling-in process feeds off the input intensity within the object interior. The reader can now better appreciate why I set C=0 to start off my exposition. Suppose that a feedforward inhibitory stage acts on an input pattern before the feedback network responds to the transformed pattern. Let the feedforward stage use its noise suppression property to convert a rectangular input pattern into an edge reaction that suppresses the rectangle's interior (Figure 11b). Then let the feedback network transform the edge-enhanced pattern. Where does the feedback network get the input energy to fill-in off the edge reactions into the pattern's interior if the interior activities have already been suppressed? How does the feedback network know that the original input pattern had an interior at all? This is the technical version of the "To Have Your Edge and Fill-In Too" dilemma that I raised in Section 17. We are now much closer to an answer.

38. The Monocular Length-Luminance Effect

Before suggesting a resolution of this dilemma, I will note a property of functional scales which seems to be reflected in various data, such as the Wallach and Adams (1954) experiment, but seems not to have been studied directly. This property concerns changes in functional scaling that are due to changes in luminance of an input pattern. To illustrate the phenomenon in its simplest form, I consider the response of a competitive feedback network such as (23) to a rectangular input pattern of increasing luminance. In Figure 12a, the rectangle intensity is too low to elicit any suprathreshold

Figure 12

reaction. In Figure 12b, a higher rectangle intensity fills-in the region with a single interior peak and two boundary peaks. At the still higher intensity of Figure 12c, two interior peaks emerge. At successively higher intensities, more peaks emerge until the intensity gets so high that a smaller number of peaks again occurs (Figure 12d). This progressive increase followed by a progressive decrease in the number of interior peaks has been found in many computer runs (Cohen and Grossberg, 1982; Ellias and Grossberg, 1975). It reflects the network's increasing sensitivity at higher input intensities until such high intensities are reached that the network starts to saturate and is gradually desensitized.

If we assume that the total area under an activity pattern within a unit spatial region estimates the lightness of the pattern, then it is tempting to interpret the above result as a perceived lightness change when the luminance of an object, but not of its background, is parametrically increased. This interpretation cannot, however, be made without extreme caution. This is true because the functional scaling change within one monocular representation may alter the ability of this representation to match the other monocular representation within a given structural scale. In other words, by replacing spatially homogeneous regions in a figure by spatially patterned functional scales, we can think about whether these patterns match or mismatch under prescribed conditions. An alteration in the scales which are capable of binocular matching implies an alteration in the scales which can energetically resonate. A complex change in per-

ceived brightness, depth, and length can hereby be caused.

Even during conditions of monocular viewing, the phenomenon depicted by Figure 1, has challenging implications. Consider an input pattern which is a figure against a ground with nonzero reflectance. Let the entire pattern be illuminated at successively higher luminances. Within the energy region of brightness constancy, the balance between the functional scales of figure and ground can be maintained. At extreme luminances, however, the sensitivity changes illustrated in Figure 12 can take effect and may cause a coordinated change in both perceived brightness and perceived length. If the functional wavelength, as opposed to a more global estimate of the total activated region within a structural scale, influences length judgements, then a small length reduction may be detectable both at low and high luminances. This effect should at the present time be thought of as an intriguing possibility rather than as a necessary prediction of the theory because, in realistic binocular networks, interactive effects between monocular and binocular cells and between multiple structural scales may alter the property of Figure 12.

39. Spreading FIRE: Pooled Binocular Edges, False Matches, Allelotropia, Binocular
    Brightness Summation, and Binocular Length Scaling

Now that the concept of a functional scale in a competitive feedback network is clearly in view, I can reintroduce the noise suppression inequalities (26) to show how the joint action of noise suppression and functional scaling can generate a filling-in resonant exchange (FIRE) that is sensitive to binocular properties such as disparity. Within the framework I have built up, starting a FIRE capable of global effects on perceived depth, form, and lightness is intuitively simple. I will nonetheless describe the main ideas in mechanistic terms, since if certain constraints are not obeyed, the FIRE will not ignite (Cohen and Grossberg, 1982). I will also restrict my attention to the simplest, or minimal, network which exhibits the properties that I seek. It will be apparent that the same types of properties can be obtained in a wide variety of related network designs.

First I will restrict attention to the case of a single structural scale, which is defined by excitatory and inhibitory kernels $D(j)$ and $E(j)$, respectively. Three main intuitions go into the construction.

Proposition I:

Only input pattern data which are spatially nonuniform with respect to a structural scale are informative (Section 18).

Proposition II:

The ease with which two monocular input patterns of fixed disparity can be binocularly fused depends on the spatial frequencies in the patterns (Sections 7 and 9). This dependence is not, however, a direct one. It is mediated by statistical preprocessing of the input patterns using nonlinear cross-correlations, as in Section 25. Henceforth when I discuss an "edge," I will mean a statistical edge rather than an edge within the input pattern itself.

Proposition III:

Filling-in a functional scale can only be achieved if there exists an input source on which the FIRE can feed (Section 33).

To fix ideas, let a rectangular input pattern idealize a preprocessed segment of a scene. The interior of the rectangle idealizes an ambiguous region and the boundaries of the rectangle idealize informative regions of the scene with respect to the structural scale in question. A copy of the rectangular input pattern is processed by each monocular representation. Since the scene is viewed from a distance, the two rectangular inputs will excite disparate positions within their respective monocular representations (Figure 13a). In general, the more peripheral boundary with respect to the foveal fixa-

Figure 13

tion point will correspond to a larger disparity.

Proposition I suggests that the rectangles are passed through a feedforward competitive network capable of noise suppression to extract their statistical edges (Figure

13b). Keep in mind that these edges are not zero-crossings. Rather their breadth is commensurate with the bandwidth of the excitatory kernel D(j) (Section 25). This property is used to realize Proposition II as follows.

Suppose that the edge-enhanced monocular patterns are matched at binocular cells, where I mean matching in the sense of Sections 22 and 24. Because these networks possess distance-dependent structural scales, the suppressive effects of mismatch are restricted to the spatial wavelength of an inhibitory scale E(j), rather than involving the entire network. Because the edges are statistically defined, the concepts of match and mismatch refer to the degree of coherence between monocular statistics, rather than to comparisons of individual edges. Three possible cases can occur.

The case of primary interest is the one in which the two monocular edge reactions overlap enough to fall within each other's excitatory on-center D(j). This will happen, for example, if the disparity between the edge centers does not exceed one-half the width of the excitatory on-center. Marr and Poggio (1979) have pointed out that, within this range, the probability of false matches is very small, in fact less than 5%. Within the zero-crossing formalism of Marr and Poggio (1979), however, the decision to restrict matches to this distance is not part of their definition of an edge. Within a theory wherein the edge computation retains its spatial scale at a topographically organized binocular matching interface, this restriction is automatic.

If this matching constraint is satisfied, then a pooled binocular edge is formed that is centered between the loci of the monocular edges (Figure 13c). See Ellias and Grossberg (1975, Figure 25) for an example of this shift phenomenon. The shift in position of a pooled binocular edge also has no analog in the Marr and Poggio (1979) theory. I suggest that this binocularly-driven shift is the basis for allelotropia (Section 10).

If the two distal edges fall outside their respective on-centers, but within their off-surrounds, then they will annihilate each other if they enjoy identical parameters, or one will suppress the other by contrast enhancement if it has a sufficient energetic advantage. This unstable competition will be used to suggest an explanation of binocular rivalry in Section 41.

Finally, the two edges might fall entirely outside each other's receptive fields. Then each edge can be registered at the binocular cells, albeit with less intensity than a pooled binocular edge, due to equations (2) and (4). A double image can then occur. I consider the dependence of intensity on matching to be the basis for binocular brightness summation (Section 13).

The net effect of the above operations is to generate two amplified pooled binocular edges at the boundaries of an ambiguous region if the spatial scale of the network can match the boundary disparities of the region. Networks which cannot make this match are energetically attenuated. Having used disparity, and thus depth, information to select suitable scales and to amplify the informative data within these scales, we must face the filling-in dilemma posed by Proposition III. How do the binocular cells know how to fill-in between the pooled binocular edges to recover a binocular representation of the entire pattern? Where do these cells get the input energy to spread the FIRE? In other words, having used noise suppression to achieve selective binocular matching, how do we bypass noise suppression to recover the form of the object?

If we restrict ourselves to the minimal solution of this problem, then one answer is strongly suggested. Signals from the pooled binocular edges are topographically fed back to the processing stage at which the rectangular input is registered. This is the stage just before the feedforward competitive step that extracts the monocular edges (Figure 14). Several important conclusions follow immediately from this suggestion:

Figure 14

1) The network becomes a feedback competitive network in which binocular matching modulates the patterning of monocular representations.

2) If filling-in can occur, a functional scale is defined within this feedback competitive network. A larger disparity between monocular patterns resonates best with a larger structural scale, which generates a larger functional scale. Thus perceived

length depends on perceived depth.

3) The activity pattern across the functional scale is constrained by the network's normalization property. Thus perceived depth influences perceived brightness, notably the lightnesses of objects which seem to lie at the same depth.

In short, if we can overcome the filling-in dilemma at all within feedback competitive shunting networks, then known dependencies between perceived depth, length, form, and lightness emerge as natural consequences. I know of no other theoretical approach in which this is true.

It remains to indicate how the FIRE can spread despite the action of the noise suppression inequalities (26). The main problem to avoid is summarized in Figure 15. Figure 15a depicts a pooled binocular edge. When this edge adds onto

Figure 15

the rectangular pattern, we find Figure 15b. Here there is a hump on the rectangle. If this pattern is then fed through the feedforward competitive network, a pattern such as that in Figure 15c is produced. In other words, the FIRE is quenched. This is because the noise suppression property of feedforward competition drives all activities outside the hump to subthreshold values before the positive feedback loops in the total network can enhance any of these activities.

I have exposed the reader to this difficulty to emphasize a crucial property of pooled binocular edges. If $C > 0$ in (24), then an inhibitory trough surrounds the edge (Figure 15d). (If C is too small to yield a significant trough, then the pooled edge must be passed through another stage of feedforward competition.) When the edge in Figure 15d is added to the rectangular input by a competitive interaction, the pattern in Figure 15e is generated. The region off the hump is no longer uniform. The uniform region is separated from the hump by a trough whose width is commensurate with the inhibitory scale $E(j)$. When this pattern is passed through the feedforward competition, Figure 15f is generated. The nonuniform region has been contrast-enhanced into a second hump, whereas the remaining uniform region has been anni-

hilated by noise suppression. Now the pattern is fed back to the rectangular pattern stage and the cycle repeats itself. A third hump is hereby generated, and the FIRE rapidly spreads, or "develops," across the entire rectangular region at a rate commensurate with the time it takes to feed a signal through the feedback loop. Since the cells which are excited by the rectangle are already processing the input pattern when the FIRE begins, it can now spread very quickly.

Some further comments need to be made to clarify how the edge in Figure 15d adds to the rectangular input pattern. The inhibited regions in the edge can generate signals only if they excite off-cells whose signals have a net inhibitory effect on the rectangle. This option is not acceptable because mismatched patterns at the binocular matching cells would then elicit FIREs via off-cell signalling. Rather, the edge activities in Figure 15d are rectified when they generate output signals. These signals are distributed by a competitive (on-center off-surround) anatomy whose net effect is to add a signal pattern of the shape in Figure 15d to the rectangular input pattern. In other words, if all signalling stages of Figure 14 are chosen competitive to overcome the noise-saturation dilemma (Section 21), then the desired pattern transformations are achieved. This hypothesis does not necessarily imply that the pathways between the processing stages are both excitatory and inhibitory. Purely excitatory pathways can activate each level's internal on-center off-surround interneurons to achieve the desired effect. From this perspective, one can see that the two monocular edge extraction stages and the binocular matching stage at the top of Figure 14 can all be lumped into a single binocular edge matching stage. If this is done, then the mechanism for generating FIREs seems elementary indeed. If competitive signalling is used to binocularly match monocular representations and to feed the results back to the monocular representations, then a filling-in reaction will spontaneously occur within the matched scales.

40. Figure-Ground Separation by Filling-In Barriers

Now that we have seen how a FIRE can spread, it remains to say how it can be

prevented from inappropriately covering the entire visual field. A case in point is the Julesz (1971) 5% solution of dots on a white background in the stereogram of Section 9. How do the different binocular disparities of the dots in the "figure" and "ground" regions impart distinct depths to the white backgrounds of these two regions? This is an issue because the same ambiguous white background fills both regions.

I suggest that the boundary disparities of the "figure" dots can form pooled binocular edges in a different spatial scale than the spatial scale that best pools binocular edges in the "ground" scale. At the binocular cells of the "ground" scale, mismatch of the monocular edges of the "figure" can produce an inhibitory trough whose breadth is commensurate with two inhibitory structural wavelengths. The spreading FIRE cannot cross a filling-in barrier (FIB) any more than a forest fire can cross a sufficiently broad trench.

Thus within a scale whose pooled binocular edges can feed off the ambiguous background activity, FIREs can spread in all directions until they run into FIBs. This mechanism does not imply that a FIRE can rush through all spaces between adjacent FIBs, because the functional scale is a coherent dynamic entity that will

collapse if the spaces between FIBs, relative to the functional scale, are sufficiently small. Thus a random placement of dots may, other things being equal, form better FIBs than a deterministic placement which permits a coherent flow of FIRE to run between rows of FIBs. A rigorous study of the interaction between (passive) texture statistics and (coherent) functional scaling may shed further light on the discriminability of figure-ground separation. The important pioneering studies of Julesz (1978) and his colleagues on texture statistics have thus far been restricted to conclusions which can be drawn from (passive) correlational estimates.

41. The Principle of Scale Equivalence and the Curvature of Activity-Scale Correlations: Fechner's Paradox, Equidistance Tendency, and Depth Without Disparity

My description of how a FIRE can be spread and blocked sheds light on several types of data from a unified perspective. Suppose, as in Section 36, that an ambiguous monocular view of an object excites all structural scales due to self-matching of the monocular data at each scale's binocular cells. Suppose that a binocular view of an object can selectively excite some structural scales more intensely than others due to the relationship between matching and activity amplification (Section 22). These assumptions are compatible with data concerning the simultaneous activation of several spatial scales at each position in the visual field during binocular viewing (Graham, Robson, and Nachmias, 1978; Robson and Graham, 1981), with data on binocular brightness summation (Blake, Sloane, and Fox, 1981; Cogan, Silverman, and Sekuler, 1981), and with data concerning the simultaneous visibility of rivalrous patterns and a depth percept (Kaufman, 1974; Kulikowski, 1978). The suggestion that a depth percept can be generated by a selective amplification of activity in some scales above others also allows us to understand why a monocular view does not lose its filling-in capability or other resonant properties, since it can excite some structural scales via self-matches; why a monocular view need no‘ have greater visual

sensitivity than a binocular view, despite the possibility of activating several scales due to self-matches, since a binocular view may excite its scales more selectively and with greater intensity due to binocular brightness summation; why a monocular view may look brighter than a binocular view (Fechner's paradox), since although the matched scales during a binocular view are amplified, so that activity that is lost by binocular mismatch in some scales is partially gained by binocular summation in other scales, the monocular view may excite more scales by self-matches; yet why a monocular view may have a more ambiguous depth than a binocular view, because a given scene may fail to selectively amplify some scales more than others due to its lack of spatial gradients (Gibson, 1950).

The selective amplification that enhances a depth percept is sometimes due to the selectivity of disparity matches, but it need not be. The experiment of Kaufman, Bacon, and Barroso (1973) shows that depth can be altered, even when no absolute disparities exist, by varying the relative brightnesses of monocular pattern features. The present framework interprets this result as an external manipulation of the energies that cause selective amplification of certain scales above others, and one that does so in such a way that the preferred scales are altered as the experimental inputs are varied.

The same ideas indicate how a combination of monocular motion cues and/or motion-dependent input energy changes can enhance a depth percept. Motions that selectively enhance delayed self-matches in certain scales above others will cause a depth percept.

The idea that depth can be controlled by the energy balance across several active scales overcomes a problem a Sperling-Dev models. Due to the competition between depth planes in these models, only one depth plane at a time can be active in each spatial location. However, there can exist only finitely many depth planes, both on general grounds due to the finite dimension of neural networks, and on specific grounds due to inferences from spatial frequency data wherein only a

few scales are needed to interpret the data (Graham, 1981; Wilson and Bergen, 1979). Why, then, don't we perceive just three or four different depths, one depth corresponding to activity in each depth plane? Why doesn't the depth seem to jump discretely from scale to scale as an object approaches us? Depth seems to change continuously as an object approaches us despite the existence of only a few structural scales. The idea that the energy balance across functional scales continuously changes as the object approaches, and thereby continuously alters the depth percept, provides an intuitively appealing answer. This idea also mechanistically explicates the popular thesis that the workings of spatial scales can be analogized to the workings of color vision, wherein the pattern of activity across a few cone receptor types forms the substrate for color percepts.

The present framework provides an explanation of Gogel's equidistance tendency (Section 4). Suppose that a monocularly viewed object of ambiguous depth is viewed which excites most, or all, of its structural scales through self-matches. Let a nearby binocularly viewed object selectively amplify the scales with which it forms the best pooled binocular edges. Let a FIRE spread with the greatest vigor through these amplified scales. When the FIRE reaches the monocular self-matches within its scale, it can amplify the activity of these self-matches, much as occurs during binocular brightness summation. This shift in the energy balance across the scales which represent the monocularly viewed object impart it with depthfulness. This conclusion follows — and this is the crucial point — even though no new disparity information is produced within the self-matches by the FIRE. Only an energy shift occurs. Thus although disparities may be sufficient to produce a depth percept, they may not be necessary to produce a depth percept.

I suggest instead that suitable correlations between activity and scaling across the network loci that represent different spatial positions produce a depth percept. Depth is perceived whenever the resonant activity distribution is "curved" among several structural scales as representational space is traversed, no matter how — monocularly or binocularly — the activity distribution achieves

its curvature. This conclusion may be restated as a deceptively simple proposition: An object in the outside world is perceived to be curved if it induces a curvature in the abstract representational space of activity-scale correlations.

Such a conclusion at first broach seems to smack of naive realism, but it is saved from the perils of naive realism by the highly nonlinear and nonlocal nature of the shunting network representation of input patterns. The conclusion does, however, provide a scientific rationale for the temptations of naive realism, and points the way to a form a neo-realism if one entertains the quantum-mechanical proposition that the curvature of an object in the outside world is also due to curved activity-scaling correlations in an abstract representational space. Such considerations lead beyond the scope of this article.

The view that all external operations that cause equivalent activity-scaling correlations generate equivalent depth percepts liberates our thinking from the current addiction to disparity computations and suggests how monocular gradients, monocular motion cues, and learned cognitive feedback signals can all contribute to a depth percept. Because of the importance of this conception to my theory, I give it a name: the principle of scale equivalence.

42. Reflectance Rivalry and Spatial Frequency Detection

The same ideas suggest an explanation of the Wallach and Adams (1954) data on rivalry between two central figures of different lightness (Section 13). Suppose that each monocular pattern generates a different functional scale when it is monocularly viewed (Section 38). Suppose, moreover, that the monocular input intensities are chosen so that the functional scales are spatially out of phase with each other. Then when a different input pattern is presented to each eye, the feedback exchange between monocular and binocular cells, being out of phase, can become rivalrous.

This explanation suggests a fascinating experimental possibility: Given an input figure of fixed size, test a series of lightness differences to the two eyes.

Can one find ranges of lightness where the functional scales are rivalrous followed by ranges of lightness in which the functional scales can match? If this is possible, then it is probably due to the fact that only certain peaks in the two scales binocularly match. The extra peaks self-match. Should this happen, it may be possible to detect small spatial periodicities in lightness such that binocular matches are brighter than self-matches. I am not certain that these differences will be visible, because the filling-in process from the locations of amplified binocular matches across the regions of monocular self-matches may totally obscure the lightness differences of the two types of matches. Such a filling-in process may be interpreted as a type of brightness summation.

Another summation phenomenon which may reflect the activation of a functional scale is the decrease in threshold contrast needed to detect an extended grating pattern as the number of cycles in the pattern is increased. Robson and Graham (1981) quantitatively explain this phenomenon "by assuming that an extended grating pattern will be detected if any of the independently perturbed detectors on whose receptive field the stimulus falls signals its presence" (p. 409). What is perplexing about this phenomenon is that "some kind of summation process takes place over at least something approaching 64 cycles of our patterns...it is stretching credulity rather far to suppose that the visual system contains detectors with receptive fields having as many as 64 pairs of excitatory and inhibitory regions" (p. 413). This phenomenon seems less paradoxical if we suppose that a single suprathreshold peak within a structural scale can drive contiguous subthreshold peaks within that scale to suprathreshold values via a disinhibitory action. Suppose moreover that increasing the number of cycles increases the expected number of suprathreshold peaks that will occur at a fixed contrast. Then a summation effect across 64 structural wavelengths is not paradoxical if it is viewed as a filling-in reaction from suprathreshold peaks to subthreshold peaks, much like the filling-in reaction that may occur between binocular matches and self-matches in the Wallach and Adams (1954) paradigm.

Due to the large number of phenomena which become intuitively more plausible using this type of filling-in idea, I believe that quantitative studies of how to vary input brightnesses to change the functional scales generated by complex visual stimuli deserve more experimental and theoretical study. One challange is to find new ways to selectively increase or decrease the activity within one structural scale without inadvertently increasing or decreasing the activities within other active scales as well. In meeting this challenge, possible effects of brightness changes on perceived length are no less interesting than their effects on perceived depth. For example, suppose that an increase in input contrast decreases the functional scale within a prescribed structural scale. Even if the individual peaks in the several functional scales retain approximately the same height, a lightness difference may occur due to the increased density of peaks within a unit cellular region. This lightness difference will alter length scaling in the limited sense that it can alter the ease with which matching can occur between monocular signals at their binocular interface, as I have just argued. It remains quite obscure, however, whether such a functional length change can also alter behavioral estimates of length, or whether behavioral length estimates are due to read-out from more global properties of the regions wherein activity is concentrated across all scales.

43. Resonance in a Feedback Dipole Field: Binocular Development and Figure-Ground Completion

My discussions of how a FIRE spreads (Section 39) and of figure-ground completion (Section 40) tacitly used properties that require another design principle to be realized. This design suggests how visual networks are organized into dipole fields consisting of subfields of on-cells and subfields of off-cells wherein the on-cells and the off-cells are joined together by a competitive interaction. Because this concept has been extensively discussed elsewhere (Grossberg, 1980; 1982 b,c), I will only sketch the properties which I need here.

I will start with a disclaimer to emphasize that I have a very specific con-
cept in mind. My dipoles are not the classical dipoles which Julesz (1971) used
to build an analog model of stereopsis. My dipoles are on-cell off-cell pairs
such that a sudden offset of a previously sustained input to the on-cell can eli-
cit a transient antagonistic rebound, or off-reaction, in the activity of the
off-cell. Similarly, a sudden and equal arousal increment to both the on-cell and
the off-cell can elicit a transient antagonistic rebound in off-cell activity if
the arousal increment occurs while the on-cell is active (Figure 16). Thus my
notion of dipole describes how STM can be rapidly reset either by temporal fluctu-

Figure 16

ations in specific visual cues, or by unexpected events, not necessarily visual
at all, which are capable of triggering an arousal increment at visually respon-
sive cells. In my theory, such an unexpected event is hypothesized to elicit the
mismatch negativity component of the N200 evoked potential, and such an antagonis-
tic rebound, or STM reset, event is hypothesized to elicit the P300 evoked poten-
tial. These reactions to specific and nonspecific inputs are suggested to be medi-
ated by slowly varying transmitter substances - notably catecholamines like norad-
renaline - which multiplicatively gate, and thereby habituate to, input signals
on their way to the on-cells and the off-cells. The outputs of these cells there-
upon compete before eliciting net on-reactions and off-reactions, respectively,
from the dipole (Figure 17).

In a dipole field, the on-cells are hypothesized to interact via a shunting

Figure 17

on-center off-surround network. The off-cells are also hypothesized to interact via
a shunting on-center off-surround network. These shunting networks normalize and
tune the STM activity within the on-subfield and the off-subfield of the total
dipole field network. The dipole interactions between on-cells and off-cells
enable an on-cell onset to cause a complementary off-cell suppression, and an on-

cell offset to cause a complementary off-cell enhancement. This duality of reac-
tions rationalizes structural neural arrangements such as on-center off-surround
networks juxtaposed against off-center on-surround networks, and a variety of
visual phenomena such as positive and negative aftereffects, the McCollough effect,
spatial frequency adaptation, monocular rivalry, and Gestalt switching between
ambiguous figures (Grossberg, 1980a).

The new features that justify mentioning dipole fields here are that the on-
fields and off-fields can interact to generate functional scales, and that the sig-
nals which regulate the balance of activity between on-cells and off-cells can habi-
tuate as the transmitter substances that gate these signals are progressively deple-
ted. These facts will now be used to clarify how figure-ground completion and bino-
cular rivalry occur. I wish to emphasize, however, that dipole fields were not inven-
ted to explain such visual effects. Rather they were invented to explain how internal
representations which self-organize (e.g., develop, learn) as a result of experience
can be stabilized against the erosive effects of later environmental fluctuations.
My adaptive resonance theory suggests how learning can occur in response to resonant
activity patterns, yet is prevented from occurring when rapid STM reset and memory
search routines are triggered by unexpected events. In the present instance, if LTM
traces are placed in the feedforward and feedback pathways that subserve binocular
resonances, then the theory suggests that binocular development will occur only in
response to resonant data patterns, notably to objects to which attention is paid
(Grossberg, 1976, 1978a, 1980a; Singer, 1982). Because the mechanistic substrates
needed for the stable self-organization of perceptual and cognitive codes are not
peculiar to visual data, one can immediately understand why so many visual effects
have analogs in other modalities.

An instructive instance of figure-ground completion is Beck's phantom letter E
(Section 7). To fully explain this percept, one needs a good model of competition
between orientation sensitive dipole fields; in particular, a good physiological
model of cortical hypercolumn organization (Hubel and Wiesel, 1977). Some observations

can be made about the relevance of dipole field organization in the absence of a complete model.

Suppose that the regularly spaced vertical dark lines of the "ground" are sufficiently dense to create a statistically smoothed pattern when they are pre-processed by the nonlinear cross-correlators of some structural scales (Glass and Switkes, 1976). When such a smoothed pattern undergoes noise suppression within a structural scale, it generates statistical edges at the boundary of the "ground" region due to the sudden change in input statistics at this boundary. These edges of the (black) off-field generate complementary edges of the (white) on-field due to dipole inhibition within this structural scale. These complementary edges can use the ambiguous (preprocessed) white as an energy source to generate a FIRE that fills in the interior of the "ground." This FIRE defines the ground as a coherent entity. The "ground" does not penetrate the "figure" because FIBs are generated by the competition which exists between orientation detectors of sufficiently different orientation.

A "figure" percept can arise in this situation as the complement of the coherently filled-in "ground", which creates a large shift in activity-scale correlations at the representational loci corresponding to the "ground" region. In order for the "figure" to achieve a unitary existence except as the complement of the "ground", a mechanism needs to operate on a broader structural scale than that of the variously oriented lines that fill the figure. For example, suppose that, due to the greater spatial extent of vertical ground lines than nonvertical figure lines, the smoothed vertical edges can almost completely inhibit all smoothed nonvertical edges near the figure-ground boundary. Then the "figure" can be completed as a disinhibitory filling-in reaction among all the smoothed nonvertical orientations of this structural scale. Thus "figure" and "ground" fill-in due to disinhibitory reactions among different subsets of cells according to this view. A lightness difference may be produced between such a "figure" and a "ground" (Dodwell, 1975).

A similar argument sharpens the description of how figure-ground completion occurs during viewing of the Julesz 5% stereogram (Section 40). In this situation, black dots that can be fused by one structural scale may nonetheless form FIBs in other structural scales. A FIRE is triggered in the structural scales with fused black dots by the disinhibitory edges which flank the dots in the scale's w.ite off-field. This FIRE propagates until it reaches FIBs that are generated by the nonfused dots corresponding to an input region of different disparity. The same thing happens in all structural scales which can fuse some of the dots. The figure-ground percept is a statistical property of all the FIREs that occur across scales.

44. Binocular Rivalry

Bi .ocular rivalry can occur in a feedback dipole field. The dynamics of a dipole field also explain why sustained monocular viewing of a scene does not routinely cause a perceived waxing and waning of the scene at the frequency of binocular rivalry, but may nonetheless cause monocular rivalry in response to

suitably constructed pictures at a rate that depends on the juxtaposition of features in the picture (Grossberg, 1980a, Section 12). Herein I will focus on how the slowly habituating transmitter gates in the dipole field can cause binocular rivalry without necessarily causing monocular waxing and waning.

Let a pair of smoothed monocular edges mismatch at the binocular matching cells. Also suppose that one edge momentarily enjoys a sufficient energetic advantage over the other edge to be amplified by contrast enhancement as the other edge is completely suppressed. This suppression can be mediated by the competition between the off-cells that correspond to the rivalrous edges. In particular, the on-cells of the enhanced edge inhibit their off-cells via dipole competition. Due to the tonic activation of off-cells, the off-cells of the other edge are disinhibited via the shunting competition that normalizes and tunes the off-field. The on-cells of these disinhibited off-cells are thereupon inhibited via dipole competition.

As this is going on, the winning edge at the binocular matching cells elicits the feedback signals that ignite whatever FIREs can be supported by the monocular data. This resonant activity gradually depletes the transmitters which gate the resonating pathways. As the habituation of transmitter progresses, the net sizes of the gated signals decrease.

The inhibited monocular representation does not suffer this disadvantage because its signals, having been suppressed, do not habituate the transmitter gates in their pathways. Finally a time may be reached when the winning monocular representation loses its competitive advantage due to progressive habituation of its transmitter gates. As soon as the binocular competition favors the other monocular representation, contrast enhancement bootstraps it into a winning position and a rivalrous cycle is initiated.

A monocularly viewed scene does not inevitably wax and wane for the following reason. Other things being equal, its transmitter gates habituate to a steady level such that the habituated gated signals are an increasing function of their input

sizes (Grossberg, 1968, 1981, 1982a). Rivalry occurs only when competitive feedback signalling, by rapidly suppressing some populations but not others, sets the stage for the competitive balance to slowly reverse as the active pathways that sustain the suppression habituate faster than the inactive pathways. The same mechanism can cause a percept of monocular rivalry to occur when the monocular input pattern contains a suitable spatial juxtaposition of mutually competitive features (Rauschecker, Campbell, and Atkinson, 1973).

45. Concluding Remarks about Filling-In and Quantization

The quantized dynamic geometry of FIRE provides a mechanistic framework in which the experimental interdependence of many visual properties may be discussed in a unified fashion. Of course, a great deal of theoretical work remains to be done, even assuming all the concepts are correct, not only in working out the physiological designs in which these dynamic transactions take place but also in subjecting the numerical and mathematical properties of these designs to a confrontation with quantitative data. Also the discussion of disinhibitory filling-in needs to be complemented by a discussion of how hierarchical feedback interactions between the feedforward adaptive filters (features) and feedback adaptive templates (expectancies) that define and stabilize a developing code can generate pattern completion effects, which are another form of filling-in (Dodwell, 1975; Grossberg, 1978a, Sections 21-22; 1980a, Section 17; Lanze, Weisstein, and Harris, 1982). Despite the incompleteness of this program, the very existence of such a quantization scheme suggests an answer to some fundamental questions.

Many scientists have, for example, realized that since the brain is a universal measurement device acting on the quantum level, its dynamics should in some sense be quantized. This article suggests a new sense in which this is true by explicating some quantized properties of binocular resonances. One can press this question further by asking why binocular resonances are nonlinear phenomena that do not take the form of classical linear quantum theory? I have elsewhere argued that this is ause of the crucial role which resonance plays in stabilizing the brain's self-organization (Grossberg, 1976, 1978a, 1980a). The traditional quantum theory is not derived from principles of self-organization, despite the fact that the evolution of physical matter is as much a fundamental problem of self-organization on the quantum level as are the problems of brain development, perception, and learning. It will be interesting to see as the years go by whether traditional quantum theory looks more like an adaptive resonance theory as it too incorporates self-organizing principles into its computational structure.

REFERENCES

Amari, S.I. and Arbib, M.A., Competition and cooperation in neural nets. In J.

    Metzler (Ed.), Systems neuroscience, New York: Academic Press, 1977.

Attneave, F., Some informational aspects of visual perception. Psychological

    Review, 1954, 61, 183-193.

Baylor, D.A. and Hodgkin, A.L., Changes in time scale and sensitivity in turtle

    photoreceptors. Journal of Physiology, 1974, 242, 729-758.

Baylor, D.A., Hodgkin, A.L. and Lamb, T.D., The electrical response of turtle

    cones to flashes and steps of light. Journal of Physiology, 1974, 242, 685-

    727(a).

Baylor, D.A., Hodgkin, A.L. and Lamb, T.D., Reconstruction of the electrical

    responses of turtle cones to flashes and steps of light. Journal of Physiology,

    1974, 242, 759-791(b).

Beck, J., Surface color perception. Ithaca, New York: Cornell University Press,

    1972.

Blake, R., Sloane, M. and Fox, R., Further developments in binocular summation.

    Perception and Psychophysics, 1981, 30, 266-276.

Blakemore, C., Carpenter, R.H. and Georgeson, M.A., Lateral inhibition between

    orientation detectors in the human visual system. Nature, 1970, 228, 37-39.

Blank, A.A., Metric geometry in human binocular perception: Theory and fact. In

    E.L.J. Leeuwenberg and H.F.J.M. Buffart (Eds.), Formal Theories of Visual

    Perception, New York: Wiley, 1978.

Caelli, T., Hoffman, W.C. and Lindman, Apparent motion: Self-excited oscillation

    induced by retarded neuronal flows. In E.L.J. Leeuwenberg and H.F.J.M. Buffart

    (Eds.), Formal Theories of Visual Perception, New York: Wiley, 1978.

Carpenter, G.A. and Grossberg, S., Adaptation and transmitter gating in vertebrate

    photoreceptors. Journal of Theoretical Neurobiology, 1981, 1, 1-42.

Cogan, A.L., Silverman, G. and Sekuler, R., Binocular summation in detection

   of contrast flashes. Perception and Psychophysics, 1982, 31, 330-338.

Cohen, M.A. and Grossberg, S., Quantitative studies of binocular resonance.

   In preparation.

Coren, S., Subjective contours and apparent depth. Psychological Review, 1972,

   79, 359-367.

Cornsweet, T.N., Visual perception. New York: Academic Press, 1970.

Day, R.H., Visual spatial illusions: A general explanation. Science, 1972, 175,

   1335-1340.

Dev, P., Perception of depth surfaces in random-dot stereograms: A neural model.

   International Journal of Man-Machine Studies, 1975, 7, 511-528.

Dodwell, P.C., Pattern and object perception. In E.C. Carterette and M.P. Fried-

   man (Eds.), Handbook of Perception, Vol. 5: Seeing. New York: Academic Press,

   1975.

Eijkman, E.G.J., Jongsma, H.J. and Vincent, J., Two-dimensional filtering, oriented

   line detectors, and figural aspects as determinants of visual illusions.

   Perception and Psychophysics, 1981, 29, 352-358.

Ellias, S.A. and Grossberg, S., Pattern formation, contrast control, and oscillations

   in the short term memory of shunting on-center off-surround networks. Biological

   Cybernetics, 1975, 20, 69-98.

Emmert, E., Grössenverhaltnisse der Nachbilder. Klin. Monatsbl. d. Augenheilk, 1881,

   19, 443-450.

Fender, D.H. and Julesz, B., Extension of Panum's fusional area in binocularly

   stabilized vision. Journal of the Optical Society of America, 1967, 57, 819-830.

Foley, J.M., Binocular distance perception. Psychological Review, 1980, 87, 411-434.

Gibson, J.J., Perception of the visual world. Boston: Houghton Mifflin, 1950.

Gilchrist, A.L., The perception of surface blacks and whites. Scientific American,

   1979, 240, 112-124.

Glass, L. and Switkes, E., Pattern recognition in humans: Correlations which cannot be perceived. Perception, 1976, 5, 67-72.

Gogel, W.C., The tendency to see objects as equidistant and its reverse relations to lateral separation. Psychological Monograph, 1956, 70 (whole no. 411).

Gogel, W.C., Equidistance tendency and its consequences. Psychological Bulletin, 1965, 64, 153-163.

Gogel, W.C., The adjacency principle and three-dimensional visual illusions. Psychono. ic Monograph Supplement, 1970, 3 (whole no. 45), 153-169.

Graham, N., The visual system does a crude Fourier analysis of patterns. In S. Grossberg (Ed.), Mathematical Psychology and Psychophysiology. Providence, R.I.: American Mathematical Society, 1981.

Graham, N, and Nachmias, J., Detection of grating patterns containing two spatial frequencies: A test of single-channel and multiple channel models. Vision Research, 1971, 11, 251-259.

Graham, N., Robson, J.G. and Nachmias, J., Grating summation in fovea and periphery. Vision Research, 1978, 18, 816-825.

Gregory, R.L., Eye and Brain. New York: McGraw-Hill, 1966.

Grossberg, S.,Some physiological and biochemical consequences of psychological postulates. Proceedings of the National Academy of Sciences, 1968, 60, 758-765.

Grossberg, S., Neural pattern discrimination. Journal of Theoretical Neurobiology, 1970, 27, 291-337.

Grossberg, S., Neural expectation: Cerebellar and retinal analogs of cells fired by learnable or unlearned pattern classes. Kybernetik, 1972, 10, 49-57.

Grossberg, S., Contour enhancement, short-term memory, and constancies in reverberating neural networks. Studies in Applied Mathematics, 1973, 52, 217-257.

Grossberg, S., Adaptive pattern classification and universal recoding, II: Feedback expectation, olfaction, and illusions. Biological Cybernetics, 1976, 23, 187-202.

Grossberg, S., A theory of human memory: Self-organization and performance of
sensory-motor codes, maps, and plans. In R. Rosen and F. Snell (Eds.),
Progress in theoretical biology, Vol. 5. New York: Academic Press: 1978(a).

Grossberg, S., Competition, decision, and consensus. *Journal of Mathematical
Analysis and Applications*, 1978, 66, 470-493(b).

Grossberg, S., Decisions, patterns, and oscillations in the dynamics of competitive
systems with applications to Volterra-Lotka systems. *Journal of Theoretical
Biology*, 1978, 101-130(c).

Grossberg, S., Behavioral contrast in short-term memory: Serial binary memory models
or parallel continuous memory models? *Journal of Mathematical Psychology*, 1978,
17, 199-219(d).

Grossberg, S., Communication, memory, and development. In R. Rosen and F. Snell
(Eds.), *Progress in theoretical biology*, Vol. 5. New York: Academic Press,
1978(e).

Grossberg, S., How does a brain build a cognitive code? *Psychological Review*, 1980,
87, 1-51(a).

Grossberg, S., Biological competition: Decision rules, pattern formation, and
oscillations. *Proceedings of the National Academy of Sciences*, 1980, 77,
2338-2342(b).

Grossberg, S., Adaptive resonance in development, perception, and cognition. In
S. Grossberg (Ed.), *Mathematical Psychology and Psychophysiology*. Providence,
R.I.: American Mathematical Society, 1981.

Grossberg, S., *Studies of mind and brain: Neural principles of learning, perception,
development, cognition, and motor control*. Boston: Reidel Press, 1982(a).

Grossberg, S., Some psychophysiological and pharmacological correlates of a
developmental, cognitive, and motivational theory. In J. Cohen, R. Karrer,
and P. Tueting (Eds.), *Cognition and brain activity*. New York: New York
Academy of Sciences, 1982(b).

Grossberg, S., The processing of expected and unexpected events during conditioning and attention: A psychophysiological theory. Psychological Review, 1982, in press. (c).

Grossberg, S. and Levine, D.S., Some developmental and attentional biases in the contrast enhancement and short term memory of recurrent neural networks. Journal of Theoretical Biology, 1975, 53, 341-380.

Grünau, M.W. von., The involvement of i-lusory contours in stroboscopic motion. Perception and Psychophysics, 1979, 25, 205-208.

Hagen, M.A. and Teghtsoonian, M., The effects of binocular and motion-generated information on the perception of depth and height. Perception and Psychophysics, 1981, 30, 257-265.

Hamada, J., A mathematical model for brightness and contour perception. Hokkaido Report of Psychology. HRP-11-76-17, 1976.

Hamada, J., Antagonistic and non-antagonistic processes in the lightness perception. Proceedings of XXIInd International Congress of Psychology, Leipzig, July 6-12, 1980.

Helmholtz, H.L.F. von., Treatise on physiological optics. J.P.C. Southall (Trans.). New York: Dover, 1962.

Hering, E., Outlines of a theory of the light sense. Cambridge, Mass.: Harvard University Press, 1964.

Holway, A.F. and Boring, E.G., Determinants of apparent visual size with distance variant. American Journal of Psychology, 1941, 54, 21-37.

Hubel, D.H. and Wiesel, T.N., Functional architecture of macaque monkey visual cortex. Proceedings of the Royal Society of London (B), 1977, 198, 1-59.

Johansson, G., About the geometry underlying spontaneous visual decoding of the optical message. In E.L.J. Leeuwenberg and H.F.J.M. Buffart (Eds.), Formal theories of visual perception. New York: Wiley, 1978.

Julesz, B., Foundations of cyclopean perception. Chicago: University of Chicago Press, 1971.

Julesz, B., Perceptual limits of texture descrimination and their implications to figure-ground separation. In E.L.J. Leeuwenberg and H.F.J.M. Buffart (Eds.), Formal theories of visual perception. New York: Wiley, 1978.

K.. marek, L.K. and Babloyantz, A., Spatiotenporal patterns in epileptic seizures. Biological Cybernetics, 1977, 26, 199-208.

Kaufman, L., Sight and mind: An introduction to visual perception. New Tork: Oxford University Press, 1974.

Kaufman, L., Bacon, J. and Barroso, F., Stereopsis without image segregation. Vision Research, 1973, 13, 137-147.

Klatt, D.H., Speech perception: A model of acoustic-phonetic analysis and lexical access. In R.A. Cole (Ed.), Perception and production of fluent speech. Hillsdale, N.J.: Erlbaum, 1980.

Koffka, K., Principles of Gestalt psychology. New York: Harcourt, Brace, 1935.

Kulikowski, J.J., Limit of single vision in stereopsis depends on contour sharpness. Nature, 1978, 275, 126-127.

Land, E.H., The retinex theory of color vision. Scientific American, 1977, 237, 108-128.

Lanze, M., Weisstein, N. and Harris, J.R., Perceived depth vs. structural relevance in the object-superiority effect. Perception and Psychophysics, 1982, 31, 376-382.

Lettvin, J.Y., "Filling out the forms": An appreciation of Hubel and Wiessel. Science, 1981, 214, 518-520.

Levelt, W.J.M., On binocular rivalry. Soesterberg (The Netherlands): Institute for Perception, RVO-TNO, 1964.

Levine, D.S. and Grossberg, S., Visual illusions in neural networks: Line neutralization, tilt aftereffect, and angle expansion. Journal of Theoretical Biology, 1976, 61, 477-504.

Luneberg, R.K., Mathematical analysis of binocular vision. Princeton, N.J.: Princeton University Press, 1947.

Marr, D. and Hildreth, E., Theory of edge detection. Proceedings of the Royal Society of London (B), 1980, 207, 187-217.

Marr, D. and Poggio, T., Cooperative computation of stereo disparity. Science, 1976, 194, 283-287.

Marr, D. and Poggio, T., A computational theory of human stereo vision. Proceedings of the Royal Society of London (B), 1979, 204, 301-328.

Newell, A., Harpy, production systems, and human cognition. In R. Cole (Ed.), Perception and production of fluent speech. Hillsdale, N.J.: Erlbaum, 1980.

O'Brien, V., Contour perception, illusion and reality. Journal of the Optical Society of America, 1958, 48, 112-119.

Osgood, C.E., Suci, G.J. and Tannenbaum, P.H., The measurement of meaning. Urbana, Ill.: University of Illinois, 1957.

Raaijmakers, J.G.W. and Shiffrin, R.M., Search of associative memory. Psychological Review, 1981, 88, 93-134.

Ratliff, F., Mach bands: Quantitative studies on neural networks in the retina. New York: Holden-Day, 1965.

Rauschecker, J.P.J., Campbell, F.W. and Atkinson, J., Colour opponent neurones in the human visual system. Nature, 1973, 245, 42-45.

Restle, F., Mathematical models in psychology. Baltimore: Penguin Books, 1971.

Richards, W., Visual space perception. In E.C. Carterette and M.P. Friedman (Eds.), Handbook of perception, Vol. 5: Seeing. New York: Academic Press, 1975.

Richards, W. and Marr, D., Computational algorithms for visual processing. M.I.T. Artificial Intelligence Lab, 1981.

Richards, W. and Miller, J.F., Jr., The corridor illusion. Perception and Psychophysics, 1971, 9, 421-423.

Robson, J.G., Receptive fields: Neural representation of the spatial and intensive attributes of the visual image. In E.C. Carterette and M.P. Friedman (Eds.), Handbook of perception (Vol. 5). New York: Academic Press, 1975.

Robson, J.G. and Graham, N., Probability summation and regional variation in contrast sensitivity across the visual field. Vision Research, 1981, 21, 409-418.

Rodieck, R.W. and Stone, J., Analysis of receptive fields of cat retinal ganglion cells. Journal of Neurophysiology, 1965, 28, 833-849.

Schwartz, E.L., Computational anatomy and functional architecture of striate cortex: A spatial mapping approach to perceptual coding. Vision Research, 1980, 20, 645-669.

Shepard, R.N., Multidimensional scaling, tree-fitting, and clustering. Science, 1980, 210, 390-398.

Shipley, T., Visual contours in homogeneous space. Science, 1965, 150, 348-350.

Singer, W. The role of attention in developmental plasticity. Human Neurobiology, 1982, 1, 41-43.

Smith A.T. and Over, R., Motion aftereffect with subjective contours. Perception and Psychophysics, 1979, 25, 95-98.

Sperling, G., Binocular vision: A physical and a neural theory. American Journal of Psychology, 1970, 83, 461-534.

Sperling, G., Mathematical models of binocular vision. In S. Grossberg (Ed.), Mathematical psychology and psychophysiology. Providence, R.I.: American Mathematical Society, 1981.

Sperling, G. and Sondhi, M.M., Model for visual luminance discrimination and flicker detection. Journal of the Optical Society of America, 1968, 58, 133-1145.

Stevens, S.S., The quantification of sensation. Daedalus, 1959, 88. 606-621.

Tschermak-Seysenegg, A. von., Introduction to physiological optics. P. Boeder (Trans.). Springfield, Ill.: C.C. Thomas, 1952.

Tynan, P. and Sekuler, R., Moving visual phantom: A new contour completion effect. Science, 1975, 188, 951-952.

Wallach, H. and Adams, P.A., Binocular rivalry of achromatic colors. American Journal of Psychology, 1954, 67, 513-516.

Watson, A.S., A Riemann geometric explanation of the visual illusions and figural after-effects. In E.C.J. Leeuwenberg and H.F.T.M. Buffart (Eds.), Formal theories of visual perception. New York: Wiley, 1978.

Weisstein, N., Maguire, W. and Berbaum, K., A phantom-motion aftereffect. Science, 1977, 198, 955-958.

Werblin, F.S., Adaptation in a vertebrate retina: Intracellular recordings in Necturus. Journal of Neurophysiology, 1971, 34, 228-241.

Werner, H., Dynamics in binocular depth perception. Psychological Monograph (whole no. 218), 1937.

Wilson, H.R. and Bergen, J.R., A four mechanism model for spatial vision. Vision Research, 1979, 19, 19-32.

FIGURE CAPTIONS

1. In (a), the luminance profile is depicted across a one-dimensional ray through the picture in (b). Although the interiors of all the regions have equal luminance, the apparent brightness of the regions is described by (c).

2. Combinations of the two pictures in (a), such as the pictures in (b), yield a depth percept when each picture is viewed through a separate eye. Depth can be seen even if the two pictures are combined to yield brightness differences but no disparity differences.

3. When the Cornsweet profile (a) and the rectangle (b) are filtered in such a way that low spatial frequencies are attenuated, both outputs look like a Cornsweet profile rather than a rectangle, as occurs during visual experience.

4. When a unit step in intensity (a) is smoothed by a Gaussian kernel, the result is (b). The first spatial derivative is (c), and the second spatial derivative is (d). The second derivative is zero at the location of the edge.

5. In this luminance profile, zero-crossings provide no information about which regions are brighter than others. Auxiliary computations are needed to determine this.

6. In the simplest feedforward competitive network, each input I_i excites its cell (population) v_i and inhibits all other populations v_j, $j \neq i$.

7. When the feedforward competitive network is exposed to the pattern in (a), it suppresses both interior and exterior regions of the pattern that look uniform to cells at these pattern locations. The result is the differential amplification of pattern regions which look nonuniform to the network, as in (b).

8. Transitions in the response of a network to a pattern (a) with multiple spatial frequencies progressively alters from (b) through (d) as the structural scales of the network expand.

9. A sigmoid signal $f(w)$ of cell activity w can suppress noise, contrast enhance suprathreshold activities, normalize total activity, and store the contrast enhanced and normalized pattern in short term memory within a suitably designed feedback competitive network.

10. In Figures 10a and 10b, the same input pattern is differently transformed and stored in STM due to different settings of the network QT.

11. Reaction of a feedforward competitive network (b) and a feedback competitive network (c) to the same input pattern (a). Only the feedback network can activate the interior of the region which receives the input pattern with unattenuated activity.

12. Response of a feedback competitive network to a rectangle of increasing luminance on a black background.

13. After the two monocular patterns (a) are passed through a feedforward competitive network to extract their nonuniform data with respect to the network's structural scales (b), the filtered patterns are topographically matched to allow pooled binocular edges to form (c) if the relationship between disparity and monocular functional scaling is favorable.

14. Monocular processing of patterns through feedforward competitive networks is followed by binocular matching of the two transformed monocular patterns. The pooled binocular edges are then fed back to both monocular representations at

a processing stage where they can feed off monocular activity to start a

FIRE.

15. The FIRE is quenched in (a) - (c) because there exists no nonuniform region

off the pooled binocular edge which can be amplified by the feedback exchange.

In (d) - (f), the inhibitory troughs of the edges enable the FIRE to propagate.

16. An antagonistic rebound, or off-reaction, in a gated dipole can be caused either

by rapid offset of a phasic input or rapid onset of a nonspecific arousal input.

As in Figure 17, function $J(t)$ represents a phasic input, function $I(t)$ repre-

sents a nonspecific arousal input, function $x_5(t)$ represents the potential, or

activity, of the on-channel's final stage, and function $x_6(t)$ represents the

potential, or activity, of the off-channel's final stage.

17. In the simplest example of a gated dipole, phasic input J and arousal input I

add in the on-channel to activate the potential x_1. The arousal input alone

activates x_2. Signals $S_1 = f(x_1)$ and $S_2 = f(x_2)$ such that $S_1 > S_2$ are hereby

generated. In the square synapses, transmitters z_1 and z_2 slowly accumulate

to a target level. Transmitter is also released at a rate proportional to

$S_1 z_1$ in the on-channel and $S_2 z_2$ in the off-channel. This is the transmitter

gating step. These signals perturb the potentials x_3 and x_4, which thereupon

compete to elicit the net on-reaction x_5 and off-reaction x_6. See Grossberg

(1980a, 1982b) for a mathematical analysis of gated dipole properties.
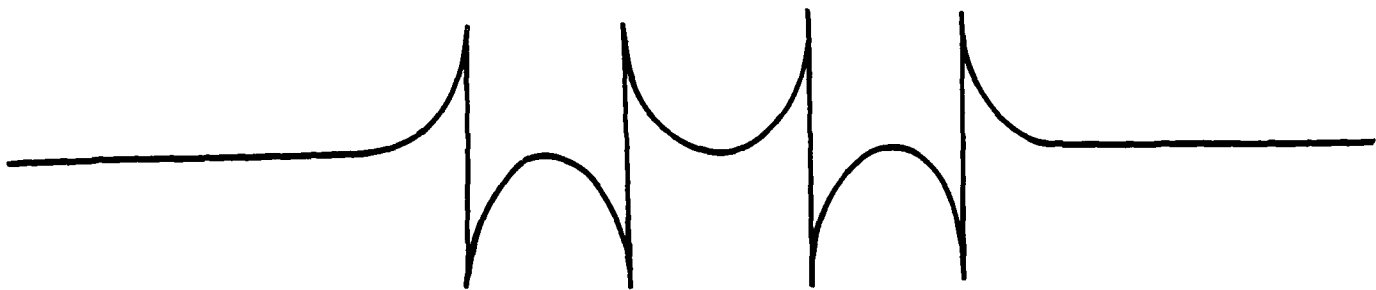
a
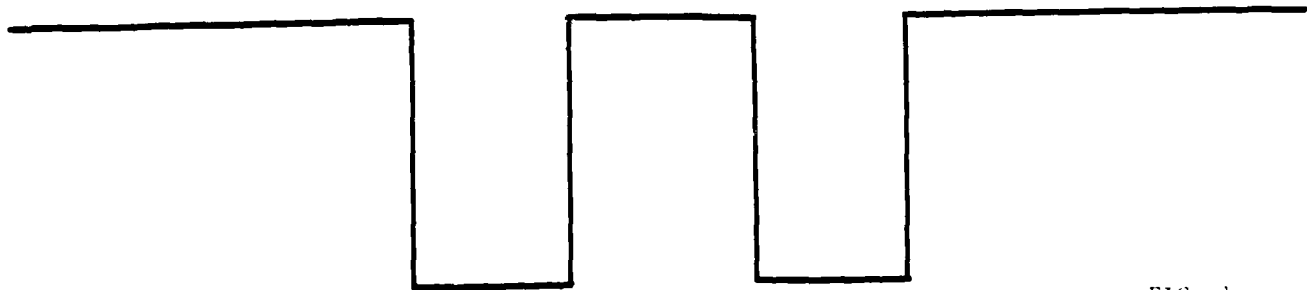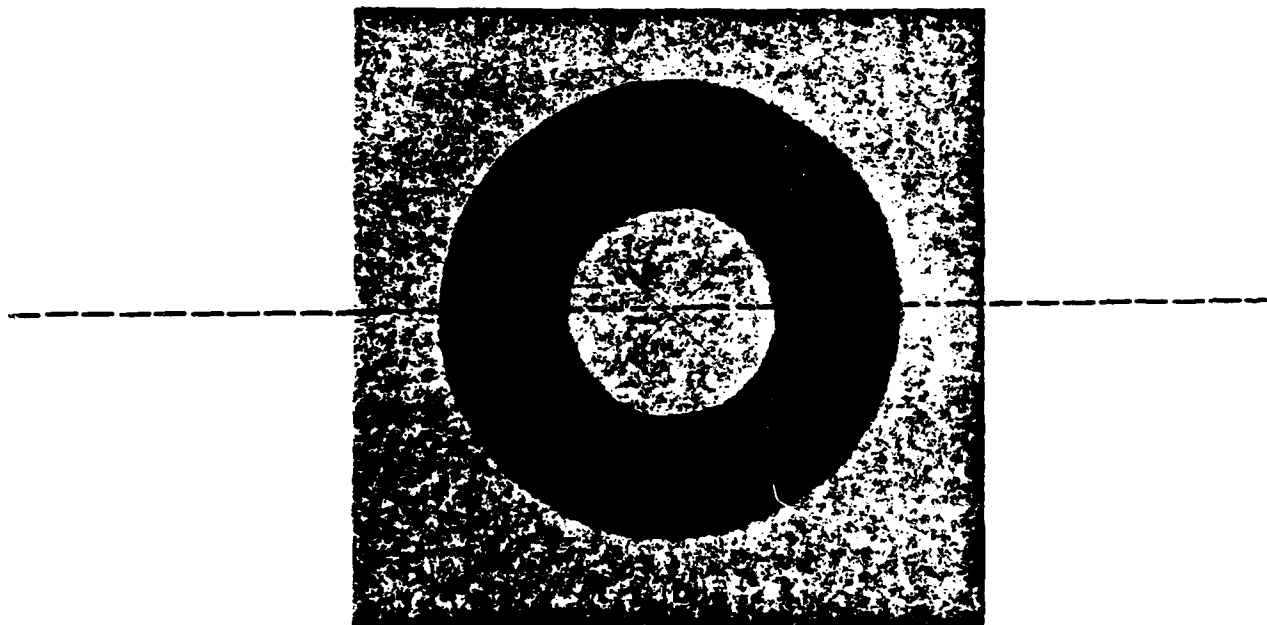
c

FIG. 1
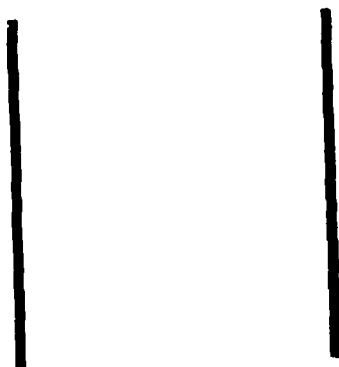
PICTURE 1   PICTURE 2

(a)

(b)

FIG. 2
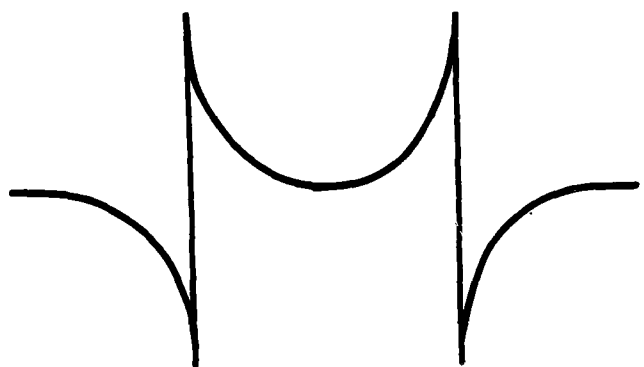
(a)
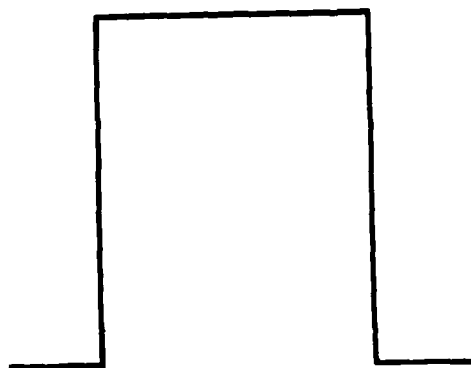
(b)

FIG. 3
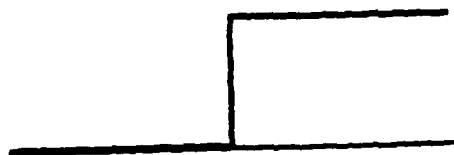
(a)

(b)

(c)
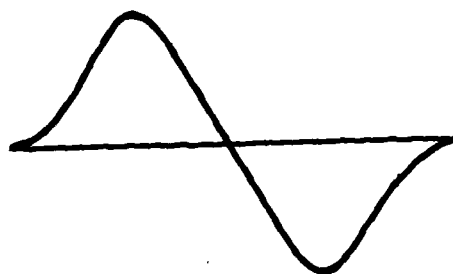
(d)

FIG. 4

FIG. 5

FIG. 6

(a)

(b)

FIG. 7

FIG. 8

SIGNAL

ACTIVITY
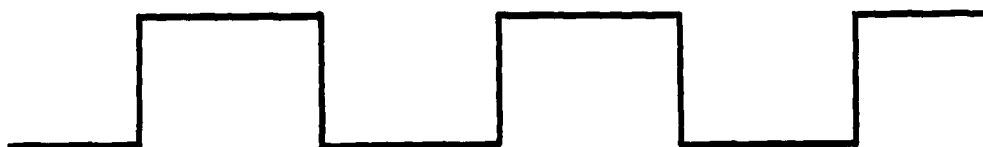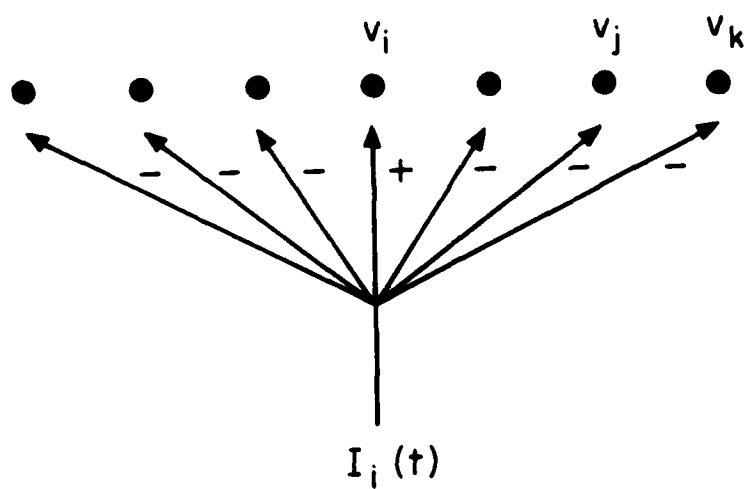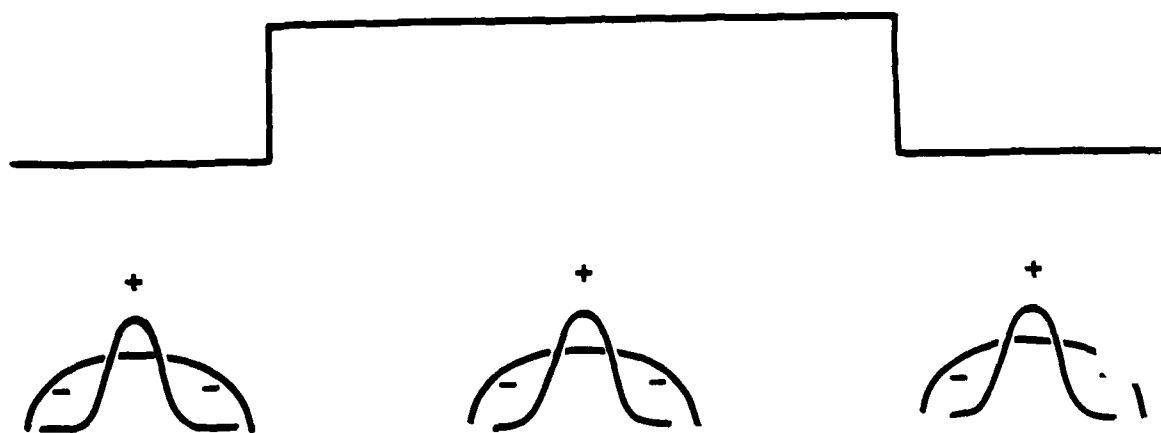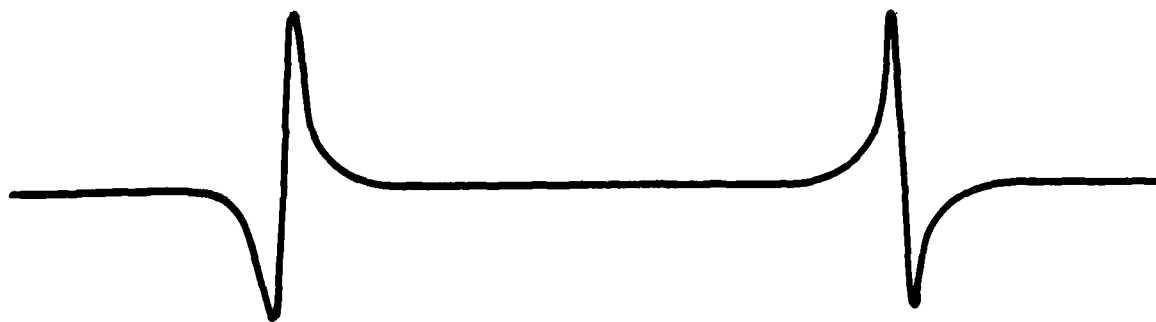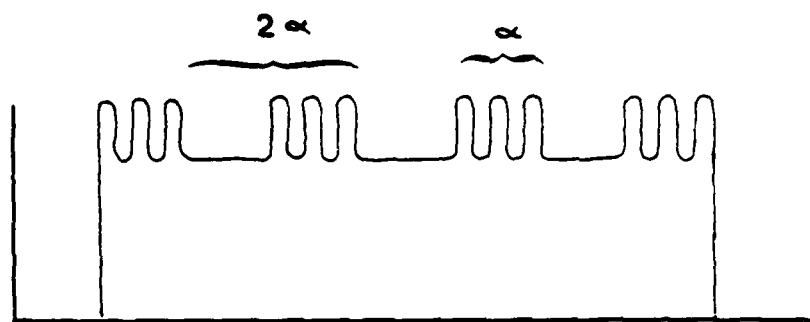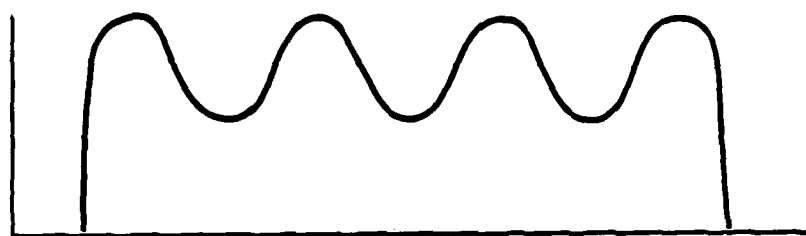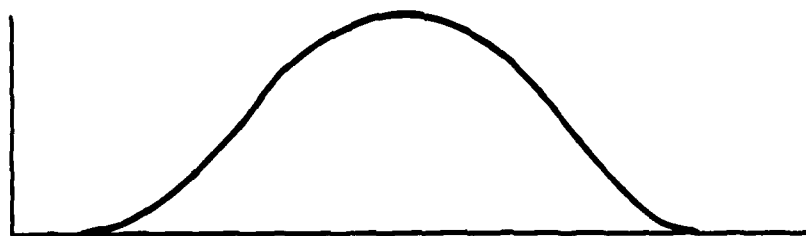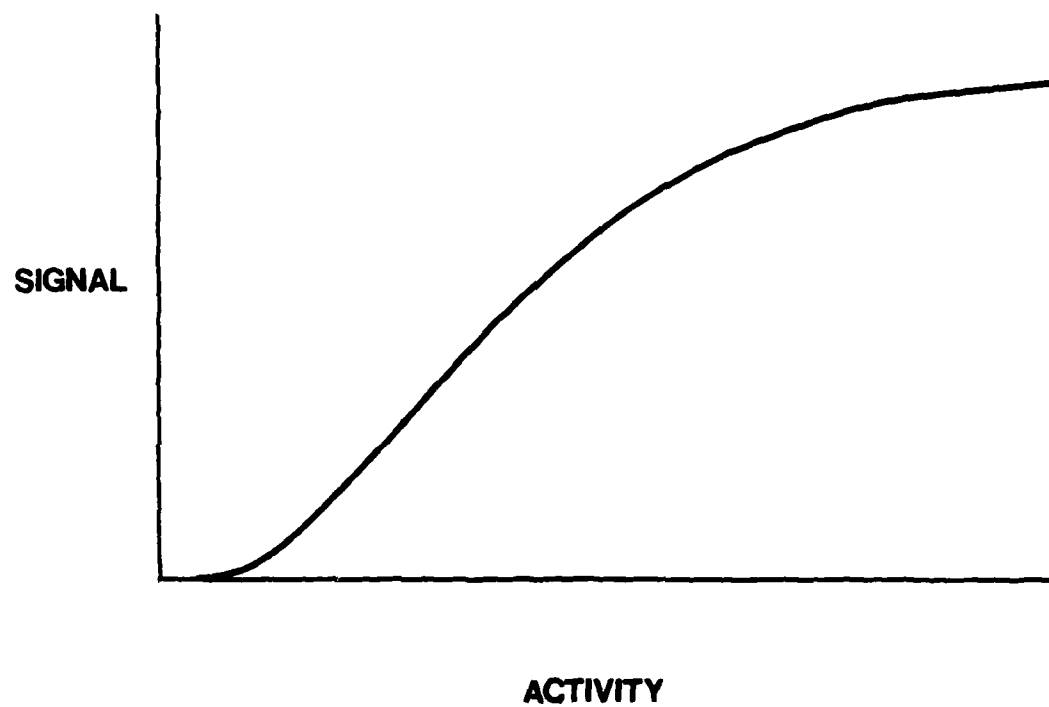
FIG. 9

PATTERN BEFORE STORAGE   PATTERN AFTER STORAGE
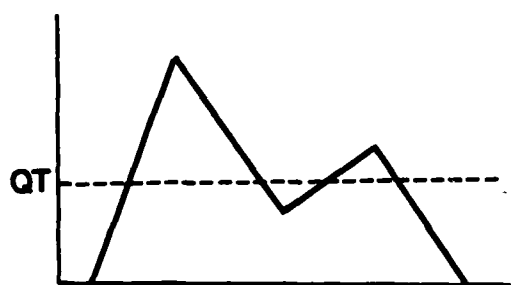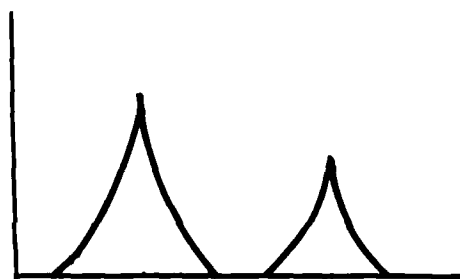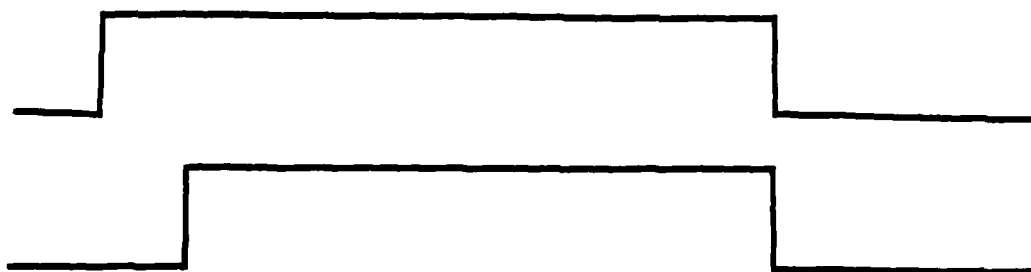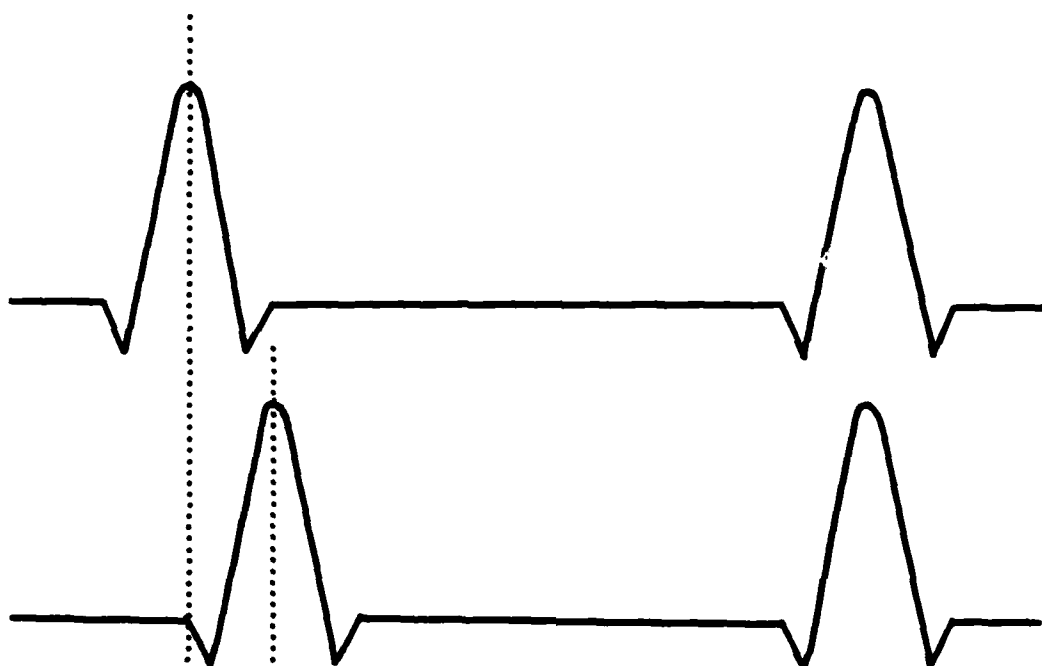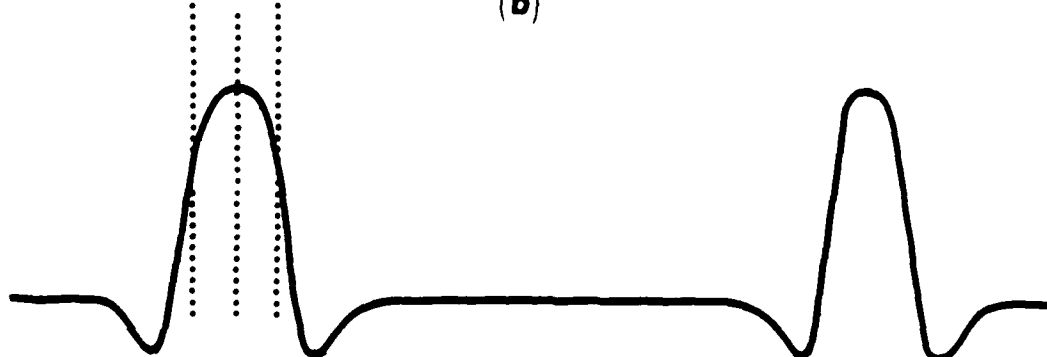

(a)
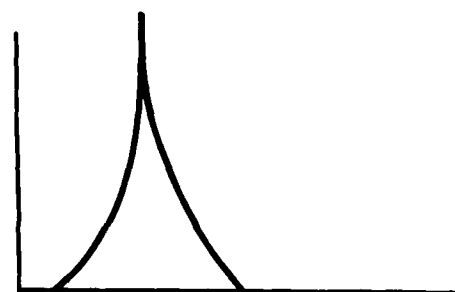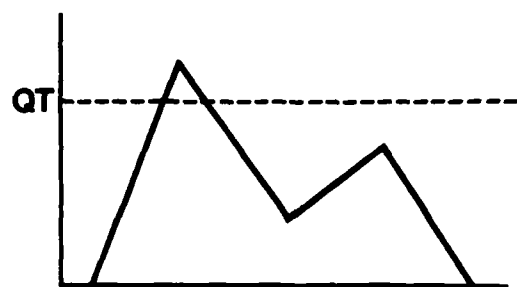
(b)

FIG. 10
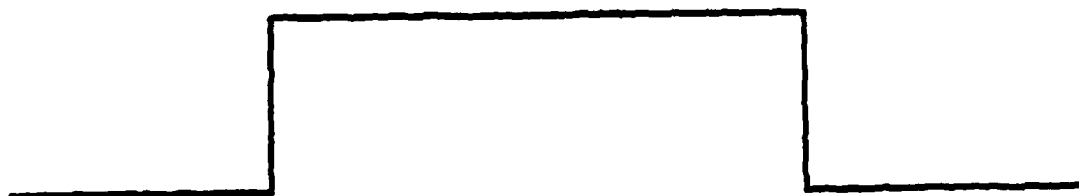
(a)
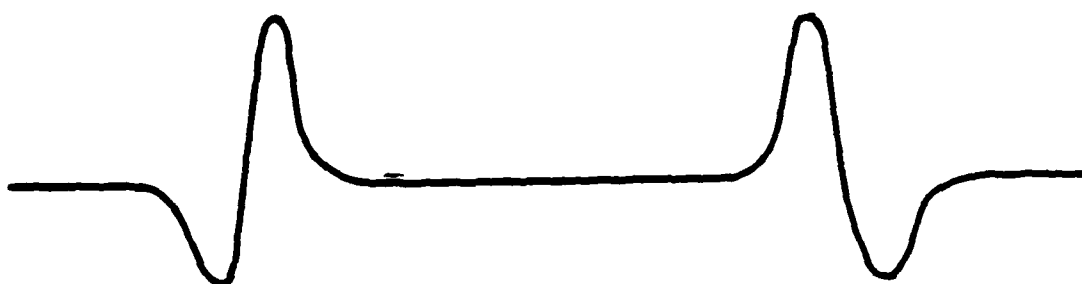
(b)
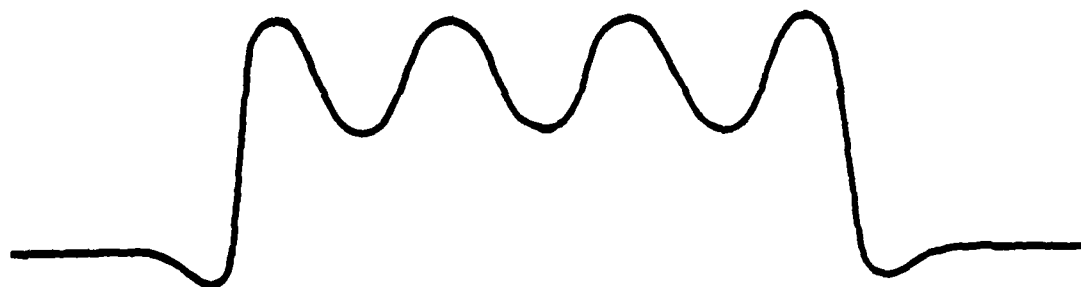
(c)

FIG. 11
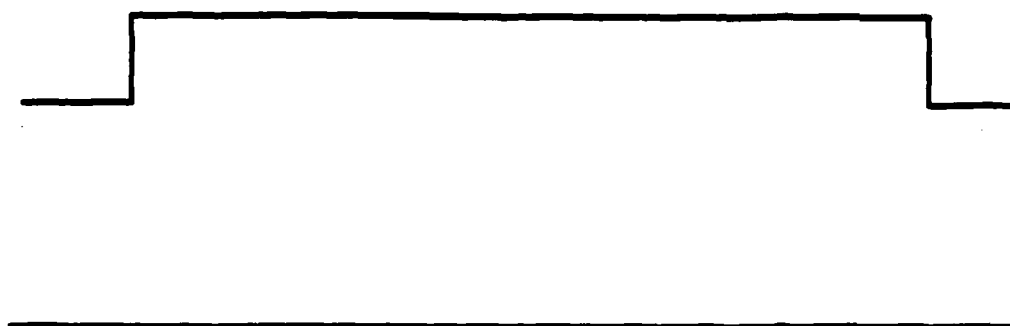
(a)
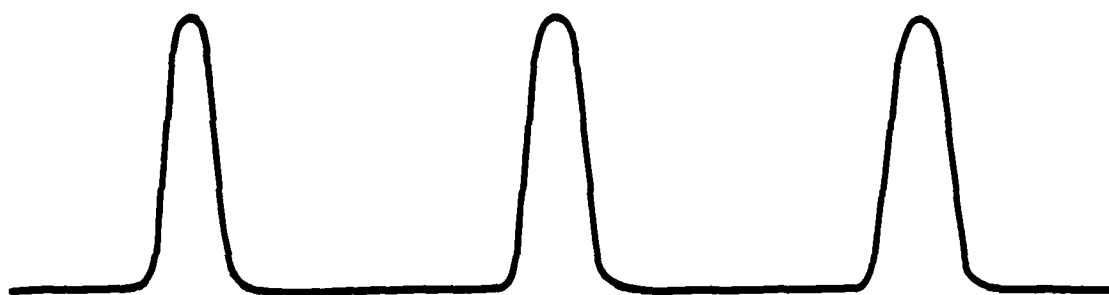
(b)

(c)

(d)

FIG. 12

(a)

(b)

(c)

FIG. 13

FIG. 14

(a)

(b)

(c)

(d)

(e)

(f)

FIG. 15

FIG. 16

ON   OFF

X_5 X_6

COMPETITION

X_3 X_4

z_1 z_2 GATE

s_1 s_2 SIGNAL

(a)

X_1 X_2

TEST   AROUSAL
INPUT   INPUT
t
(J)   (I)

FIG. 17